



FACULTÉ DES SCIENCES
DÉPARTEMENT DE MATHÉMATIQUE

Arbres et bases rationnelles

Mémoire de fin d'études présenté en vue de l'obtention du titre de
Master en Sciences Mathématiques, à finalité approfondie

Année académique 2022–2023

Auteur :
Pauline HREBENAR

Promoteur :
Michel RIGO

Remerciements

J'aimerais commencer par remercier mon promoteur, Michel Rigo, pour ses nombreuses relectures attentives et pour le temps qu'il m'a accordé durant la rédaction de ce travail.

Je remercie également mes amies de toujours Sarah Lodomez et Maria Remouchamps, ma soeur Julie, ainsi que Nicolas Delbovier pour leur soutien moral indéfectible durant ces 5 années.

Finalement, j'aimerais aussi remercier tous les professeurs que j'ai eu la chance de croiser pendant mes études et qui ont plus que jamais réussi à me faire apprécier cette magnifique matière que sont les mathématiques.

Table des matières

1	Rappels	5
1.1	Mots et langages	5
1.2	Automate fini déterministe avec sortie	6
1.2.1	Autres automates	7
1.3	Graphes et arbres	9
1.3.1	Arbre lexicographique	10
1.3.2	Arbre associé à un langage clos par préfixe	13
1.3.3	Arbre associé à une signature purement périodique labellisée	14
2	Système de numération	17
2.1	Systèmes de numération abstraits	17
2.2	Base rationnelle	18
2.2.1	Fonction d'évaluation	18
2.2.2	Représentation en base $\frac{p}{q}$	19
2.3	Systèmes de numération positionnels	23
3	Suites $\frac{p}{q}$-automatiques	27
3.1	Généralisations d'un théorème de Cobham	27
3.1.1	Cas de systèmes de numération abstraits	28
3.1.2	Cas des SNA sur un langage non régulier	33
3.2	Caractérisation des suites \mathcal{S} -automatiques	45
3.2.1	Caractérisation des suites $\frac{p}{q}$ -automatiques	49
4	Suites $\frac{p}{q}$-régulières	60
4.1	Suites k -régulières	60
4.2	Suites \mathcal{S} -régulières	63
4.2.1	Caractérisation des suites \mathcal{S} -régulières en terme d'arbre	67
4.3	Suites $\frac{p}{q}$ -régulières	74
4.3.1	Propriétés classiques conservées	77
4.3.2	Comment montrer qu'un arbre est (\mathbb{L}, h) -linéaire?	80

Introduction

Depuis le 20^e siècle, la branche des mathématiques appelée combinatoire des mots prend de plus en plus d'ampleur grâce notamment aux travaux du mathématicien Alex Thue [27]. La combinatoire des mots étudie, comme son nom l'indique, les mots (finis ou infinis) qui sont en réalité une suite de symboles appartenant à un alphabet. Un des premiers mots infinis à avoir été étudié est le mot de Thue–Morse $\mathbf{t} = 01101001 \dots$ qui est le point fixe commençant par 0 du morphisme suivant

$$\sigma : \begin{cases} 0 & \mapsto 01 \\ 1 & \mapsto 10. \end{cases}$$

Ce mot est ce qu'on appelle un mot sans cube, c'est-à-dire qu'il n'existe pas de mot fini non vide w sur l'alphabet $\{0, 1\}$ tel que www est un facteur de \mathbf{t} . Il a également la particularité d'être 2-automatique, ce qui veut dire qu'il existe un automate \mathbb{A} (appelé automate fini déterministe avec sortie) qui l'engendre. Plus précisément, l'automate \mathbb{A} est comparable à une machine prenant en argument un mot fini particulier sur l'alphabet $\{0, 1\}$ et produisant une lettre en fonction de ce mot. L'automate \mathbb{A} a dans notre cas la propriété de produire t_n (la n -ième lettre de \mathbf{t}) lors de la lecture de la représentation en base 2 de n , et ce pour tous les n entiers naturels. Dans son article [10], Alan Cobham démontre que les mots k -automatiques sont précisément ceux qui sont points fixes d'un morphisme k -uniforme (le morphisme σ est 2-uniforme car les images de toutes les lettres sont de longueur 2) par un codage (qui est un morphisme 1-uniforme).

Ce travail se base principalement sur les articles [25] et [26] et a entre autre pour vocation de généraliser ce théorème de Cobham à d'autres systèmes de numération, comme les systèmes de numération abstraits et les systèmes de numération en bases rationnelles, en donnant en argument à l'automate des mots qui sont la représentation des naturels dans ces systèmes de numération. Ils sont respectivement appelés mots \mathcal{S} -automatiques et $\frac{p}{q}$ -automatiques. Les mots automatiques sont des mots ayant un alphabet fini, il en existe donc une généralisation à un alphabet infini appelés mots k -réguliers. Nous essayons dans leur cas aussi de définir une notion de mots $\frac{p}{q}$ -réguliers et \mathcal{S} -réguliers.

Dans le premier chapitre, nous commençons par rappeler toutes les notions de bases liées à la combinatoire des mots (mots, langages, alphabets,...) ainsi que celles liées aux automates. En particulier, nous introduisons les automates finis déterministes avec sortie qui engendrent les mots dit automatiques. Nous rappelons également quelques notions de théorie des graphes, et plus précisément de la façon dont un arbre labellisé peut être associé à un langage clos par préfixe. Nous abordons également le lien entre l'arbre lexicographique d'un langage L et la régularité de celui-ci.

Par la suite, nous introduisons dans la chapitre 2 les systèmes de numération nécessaire à la définition des suites automatiques. Nous introduisons en particulier les systèmes de numération en bases entières et en bases rationnelles, ainsi que des généralisations de ceux-ci appelées systèmes de numération abstraits (notés SNA) et systèmes de numération positionnels.

Dans le chapitre 3, nous généralisons le théorème de Cobham de [10] à d'autres systèmes de numération. Nous commençons par le généraliser aux systèmes de numération abstraits sur un langage régulier. Les suites $\frac{p}{q}$ -automatiques étant des suites \mathcal{S} -automatiques pour un SNA $\mathcal{S} = (L_{\frac{p}{q}}, \mathcal{A}_p, <)$ où le langage $L_{\frac{p}{q}}$ n'est pas régulier, la deuxième partie du chapitre permet d'obtenir un théorème valable dans le cas des suites \mathcal{S} -automatiques pour un SNA \mathcal{S} construit sur un langage associé à un arbre ayant une signature labellisée purement périodique. Par la suite, nous caractérisons les suites $\frac{p}{q}$ -automatiques en termes d'arbres ayant un nombre fini de suffixes (appelés arbres rationnels).

Finalement, nous généralisons dans le chapitre 4 la notion de mots k -automatiques et $\frac{p}{q}$ -automatiques à un alphabet infini. Ces mots sont dits k -réguliers et $\frac{p}{q}$ -réguliers. Pour ce faire, nous introduisons la notion de k -noyau d'une suite nous permettant de définir de façon similaire le \mathcal{S} -noyau d'une suite dans le cas d'un SNA associé à un langage régulier. Comme le langage $L_{\frac{p}{q}}$ n'est pas régulier, nous introduisons des arbres h -linéaires afin de caractériser les suites k -régulières et \mathcal{S} -régulières (pour un SNA \mathcal{S} sur un langage régulier) pour définir par la suite les mots $\frac{p}{q}$ -réguliers.

Chapitre 1

Rappels

Ce premier chapitre a pour vocation d'introduire toutes les notions de bases nécessaires à la compréhension de ce travail. Nous commençons par rappeler des définitions et propriétés de bases de théorie des langages. Ensuite, nous rappelons la définition d'automate et plus précisément celle d'automate fini déterministe avec sortie. De là, quelques définitions de théorie de graphes sont rappelées, afin de faire le lien entre les graphes particuliers appelés arbres et certains langages, appelés langages clos par préfixe.

1.1 Mots et langages

- Un *alphabet* \mathcal{A} est un ensemble fini de lettres. Pour tout $p \in \mathbb{N} \setminus \{0\}$, nous notons $\mathcal{A}_p = \{0, \dots, p-1\}$.
- Un *mot* sur \mathcal{A} est une suite finie ou infinie de symboles. Nous notons \mathcal{A}^* l'ensemble des mots finis sur l'alphabet \mathcal{A} et $\mathcal{A}^{\mathbb{N}}$ l'ensemble des mots infinis sur ce même alphabet. Un *mot infini* \mathbf{x} sur un alphabet \mathcal{A} est une application

$$\mathbf{x} : \mathbb{N} \rightarrow \mathcal{A}$$

qui est noté $\mathbf{x} = x_0x_1x_2 \dots$.

Pour $w = w_0w_1w_2 \dots$ la *i*-ème lettre d'un mot w est notée w_i .

- La *longueur* d'un mot fini w est noté $|w|$, et compte le nombre de symboles de w . Le seul mot de longueur 0 est ε , appelé le mot vide.
- Si $a \in \mathcal{A}$ et $w = w_0 \dots w_k$ un mot de \mathcal{A}^* alors nous pouvons définir

$$|w|_a = \#\{i \in \{0, \dots, k\} \mid w_i = a\}.$$

- Soit $w = w_0 \dots w_k \in \mathcal{A}^*$. Les *préfixes* de w sont les mots de l'ensemble

$$\text{Pref}(w) = \{w_0 \dots w_j \mid j \in \{0, \dots, k\}\} \cup \{\varepsilon\}.$$

Pour un mot infini $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$, l'ensemble des préfixes de \mathbf{x} est défini comme

$$\text{Pref}(\mathbf{x}) = \{x_0 \cdots x_j \mid j \in \mathbb{N}\} \cup \{\varepsilon\}.$$

De la même façon, les *suffixes* de w et les *facteurs* de w sont les mots des ensembles suivants

$$\text{Suff}(w) = \{w_j \cdots w_k \mid j \in \{0, \dots, k\}\} \cup \{\varepsilon\},$$

$$\text{Fact}(w) = \{w_i \cdots w_j \mid i, j \in \{0, \dots, k\}, i \leq j\} \cup \{\varepsilon\}.$$

Le facteur $w_i \cdots w_j$ de w sera noté $w[i, j]$. La notion de suffixe n'existe pas dans le cas des mots infinis et les facteurs d'un tel mot sont définis comme dans le cas des mots finis, en laissant varier $i \leq j$ dans \mathbb{N} .

À partir des mots que nous venons de définir, nous pouvons introduire des ensembles de mots appelés *langages*.

Définition 1.1.1. Un *langage* L sur \mathcal{A} est un sous ensemble de \mathcal{A}^* . Le langage vide est noté \emptyset . De plus, pour tout $w \in \mathcal{A}^*$ nous pouvons définir le langage $w^{-1}L$ tel que

$$w^{-1}L = \{u \in \mathcal{A}^* \mid wu \in L\}.$$

Si \mathcal{A} est un alphabet de chiffres, nous notons également $\text{inc}(\mathcal{A}^*)$ le langage défini comme

$$\text{inc}(\mathcal{A}^*) = \{a_0 \cdots a_n \in \mathcal{A}^* \mid a_0 < a_1 < \cdots < a_n\}.$$

Certains langages vont particulièrement nous intéresser. Commençons par les *langages clos par préfixe*.

Définition 1.1.2. Un langage L composé de mots finis ou infinis sur l'alphabet \mathcal{A} est *clos par préfixe* si pour tout $w \in L$, on a $\text{Pref}(w) \in L$.

Définition 1.1.3. L'ensemble des *langages réguliers* sur l'alphabet \mathcal{A} est la plus petite famille de langages contenant le langage vide, les langages $\{a\}$ réduits à une lettre (où $a \in \mathcal{A}$) et qui est stable pour les opérations d'union, de concaténation et d'étoile de Kleen.

1.2 Automate fini déterministe avec sortie

Certains automates peuvent produire une lettre lorsqu'ils lisent un mot. Ces automates sont appelés AFDS.

Définition 1.2.1. Un *automate fini déterministe avec sortie* (AFDS) est un 6-uple $\mathbb{A} = (Q, q_0, \tau, \mathcal{A}, \delta, \mathcal{B})$ où

- Q est un ensemble fini dont les éléments sont les états de \mathbb{A}
- $q_0 \in Q$ est un état privilégié appelé état initial
- \mathcal{A} est l'alphabet de l'automate
- $\delta : Q \times \mathcal{A} \rightarrow Q$ est la fonction de transition de \mathbb{A} . Nous supposons que la fonction δ est totale, i.e., que δ est définie pour tout couple $(q, \sigma) \in Q \times \mathcal{A}$.
- $\tau : Q \rightarrow \mathcal{B}$ est appelée *fonction de sortie*. Cette fonction permet de produire une lettre lors de la lecture d'un mot dans l'AFDS. L'alphabet \mathcal{B} est appelé *alphabet de sortie*.

Exemple 1.2.2. Soit l'AFDS tel que $Q = \{A, B, C\}$, $q_0 = A$, $\mathcal{A} = \{0, 1\}$, $\mathcal{B} = \{0, 1\}$ et de fonction de sortie $\tau(A) = 0$, $\tau(B) = 1$ et $\tau(C) = 1$, représenté à la Figure 1.1. La lecture de $110 \in \mathcal{A}^*$ dans l'AFDS commence dans l'état A et parcourt successivement les états $A - B - C$ pour s'achever dans l'état C . La fonction de sortie τ donne donc la sortie 1 pour l'entrée correspondant au mot 110.

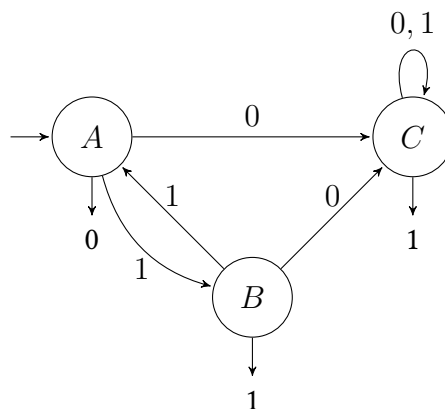


FIGURE 1.1 – Un premier AFDS.

1.2.1 Autres automates

Les *automates finis déterministes* (noté AFD) sont des cas particuliers d'AFDS avec une fonction de sortie renvoyant un booléen (1 ou 0). Les états pour lesquels la fonction de sortie rend 1 sont dit *accepteurs* (ou finaux). L'ensemble des mots qui, quand ils sont lus dans l'AFD, mènent à un état accepteur sont les mots *acceptés* par l'automate et forment le *langage accepté* par celui-ci. Le langage accepté est noté

$$\mathcal{L}(\mathbb{A}) = \{w \in \mathcal{A}^* \mid \tau(\delta(q_0, w)) = 1\}.$$

Exemple 1.2.3. Le langage accepté par l'AFD construit sur l'AFDS de la Figure 3.1 est l'ensemble des mots $w \in \{0, 1\}^*$ dont le nombre de 1 est impair.

Un *automate fini non déterministe*, noté AFND, est une généralisation d'un AFD où le nombre d'états initiaux peut-être supérieur à 1 (cet ensemble est alors noté I) et où la fonction de transition δ devient une relation de transition $\Delta \subseteq Q \times \mathcal{A}^* \times Q$. Un mot est accepté par un AFND s'il existe au moins un chemin dans l'automate menant à un état accepteur.

Remarque 1.2.4. 1) L'ensemble des états finaux d'un AFD ou d'un AFND peuvent être notés F . Dans ce cas, ils sont respectivement dénotés par $\mathbb{A} = (Q, q_0, \mathcal{A}, \delta, F)$ et $\mathbb{A} = (Q, I, \mathcal{A}, \Delta, F)$.

2) Dans un AFD, la lecture d'une lettre a depuis un certain état mène dans un unique état, i.e., deux transitions sortantes d'un même état ne peuvent pas être labellisées par la même lettre.

Les langages acceptés par les AFD et les AFND sont exactement une catégorie de langages introduite précédemment.

Théorème 1.2.5 (Kleen). *Un langage est régulier si et seulement si il est accepté par un automate fini.*

Remarque 1.2.6. Le fait que l'automate soit déterministe ou non dans le Théorème 1.2.5 n'a pas d'importance. Tout automate fini non déterministe peut être rendu déterministe grâce à la *construction par sous-ensemble*.

Tout langage régulier L étant accepté par automate fini, nous pouvons donc nous demander s'il existe un automate fini particulier acceptant le langage ayant moins d'états que les autres. Cet automate fini existe et est appelé *automate minimal* associé au langage L . Il est noté \mathbb{A}_L et est unique à isomorphisme près. Par définition de l'automate minimal, chacun de ses états correspond à un langage de la forme $u^{-1}L$ pour $u \in \mathcal{A}^*$ et l'état initial correspond en particulier au langage $\varepsilon^{-1}L$. De plus, la fonction de transition δ_L de l'automate \mathbb{A}_L est définie telle que $\delta_L(w^{-1}L, u) = (wu)^{-1}L$ pour tous mots finis w, u sur l'alphabet de l'automate \mathbb{A}_L .

Nous obtenons grâce à cet automate une autre caractérisation des langages réguliers.

Théorème 1.2.7. *Un langage L est régulier si et seulement si l'automate minimal \mathbb{A}_L a un nombre fini d'états.*

Une congruence \sim_L , appelée *congruence de Nérode*, peut être définie sur \mathcal{A}^* tel que $L \subseteq \mathcal{A}^*$. Pour tout $x, y \in \mathcal{A}^*$, nous avons $x \sim_L y$ si et seulement si $x^{-1}L = y^{-1}L$. Un langage $L \subseteq \mathcal{A}^*$ est donc régulier si et seulement si le nombre de classe d'équivalence de la congruence de Nérode sur \mathcal{A}^* est finie.

1.3 Graphes et arbres

Commençons par introduire le concept de *graphe non orienté*.

Définition 1.3.1. Soient V un ensemble et E une partie de $V \times V$. Le *graphe* $G = (V, E)$ est la donnée du couple (V, E) . Les éléments de V sont appelés les *noeuds* et les éléments de E sont appelés *arêtes*.

Les graphes non orientés qui vont nous intéresser sont appelés *arbres*.

Définition 1.3.2. Un *arbre* est un graphe non orienté, connexe et sans cycle.

Parmi tous les noeuds de l'arbre, nous pouvons en choisir un v_0 qui est privilégié. Nous notons v_0 la *racine* de l'arbre et l'arbre est alors dans ce cas un *arbre pointé*.

Les autres noeuds de l'arbre peuvent être classés par niveau. Si nous définissons la distance $\text{dist}(\cdot, \cdot)$ entre deux noeuds de l'arbre comme le nombre d'arêtes les séparant¹, alors le deuxième niveau correspond à l'ensemble des $v \in V$ tels que $\text{dist}(v_0, v) = 2$. La racine est donc l'unique noeud de niveau 0. Les *feuilles* de l'arbre sont les noeuds de l'arbre ayant uniquement pour noeuds adjacents des sommets du niveau supérieur. Les *fil*s d'un noeud $v \in V$ de niveau i sont les noeuds $u \in V$ de niveau $i + 1$ vérifiant $\text{dist}(v, u) = 1$. Inversement, v est appelé le *père* de u . Un arbre dont tous les noeuds ont k fils est appelé *arbre k -aire* complet.

Soit $T = (V, E)$ un arbre et supposons que les fils $v_{i,1}, \dots, v_{i,l}$ d'un noeud $v_i \in V$ sont ordonnés par un ordre connu et fixé. Plusieurs méthodes permettant de parcourir l'ensemble des noeuds d'un arbre existent. Le *parcours en largeur* énumère les noeuds de l'arbre niveau par niveau de gauche à droite. Ce parcours est particulièrement intéressant car il convient aussi bien pour les *arbres finis* que pour les *arbres infinis*². D'autres parcours en profondeur existent, tels que le *parcours infixe*, *parcours suffixe* et le *parcours préfixe* qui ne seront pas utilisés.

Remarque 1.3.3. Les méthodes de parcours infixe, suffixe et préfixe ne conviennent pas pour l'énumération des noeuds d'un arbre infini, car certains noeuds de l'arbre ne sont jamais parcourus.

Lorsqu'un parcours d'arbre a été sélectionné, les noeuds de l'arbre considéré peuvent être ordonnés en les numérotant grâce à l'énumération fournie par celui-ci. Dans ce cas, l'arbre est qualifié d'*arbre ordonné*. Dans la suite, tout arbre ordonné est ordonné grâce au parcours en largeur.

1. Dans le cas où il n'y a pas unicité du chemin entre deux noeuds, la distance se définit comme étant le nombre d'arêtes composant le plus court chemin entre eux-ci. Cela ne pose donc pas problème pour les arbres étant donné qu'ils n'ont pas de cycles par définition.

2. En effet de cette façon, nous sommes certains d'énumérer tous les sommets, à condition de se limiter à des sommets de degré fini.

Exemple 1.3.4. L'arbre représenté à la Figure 1.2 est un arbre pointé de racine v_0 . Les noeuds sont ordonnés grâce au parcours en largeur.

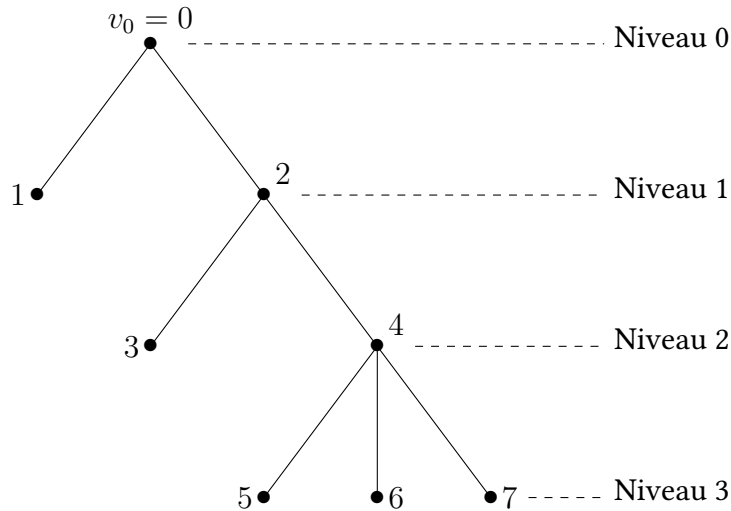


FIGURE 1.2 – Un arbre pointé ordonné de racine v_0 .

1.3.1 Arbre lexicographique

Pour décorer un arbre, nous pouvons l'agrémenter de labels sur ses noeuds et sur ses arêtes.

Définition 1.3.5. Soit $T = (V, E)$ un arbre.

- T est *labellisé* par l s'il existe une application

$$l : E \rightarrow B$$

où B est un ensemble quelconque. Pour tout $e \in E$, nous appelons $l(e)$ le label de e . Si $e = (e_1, \dots, e_p) \in E^p$ est un chemin de T , alors le label du chemin est la concaténation des labels des arêtes le composant, i.e., $l(e) = l(e_1) \cdots l(e_p)$.

- Une *décoration* de T est une application d telle que

$$d : V \rightarrow B$$

où B est un ensemble quelconque. Pour tout $v \in V$, nous appelons $d(v)$ la décoration de v .

Dans la définition précédente, B peut-être un alphabet fini. Dans ce cas, un mot sur B peut labelliser ou décorer un arbre.

Considérons $L \subseteq \mathcal{A}^*$ un langage sur un alphabet composé de $k \geq 1$ éléments. Nous pouvons construire un arbre \mathcal{T}_L dont l'ensemble des noeuds est en bijection avec \mathcal{A}^* , appelé *arbre lexicographique* associé à L . Chaque noeud de l'arbre à k arêtes de sortantes labellisées par les lettres de \mathcal{A} , ce qui fait de l'arbre \mathcal{T}_L un arbre k -aire complet. En parcourant un chemin e dans l'arbre depuis la racine, nous lisons le mot $l(e) \in \mathcal{A}^*$. Les noeuds de l'arbre \mathcal{T}_L étant en bijection avec les mots de \mathcal{A}^* , le noeud atteint dans l'arbre en lisant le mot $v \in \mathcal{A}^*$ est également appelé v . En particulier, la racine de \mathcal{T}_L est ε . Si v est un mot de L , alors $d(v) = 1$. Dans le cas contraire, $d(v) = 0$.

Exemple 1.3.6. Considérons le langage $L \subseteq \{a, b\}^*$ défini comme

$$L = \{a^n b^m \mid n, m \in \mathbb{N}\}.$$

L'arbre lexicographique associé à ce langage est un arbre binaire complet et est représenté à la Figure 1.3. Dans cet arbre, les noeuds v tels que $d(v) = 1$ sont colorés en noir. Les noeuds v tels que $d(v) = 0$ sont colorés en rouge. Cet arbre est infini car le langage $\{a, b\}^*$ est infini.

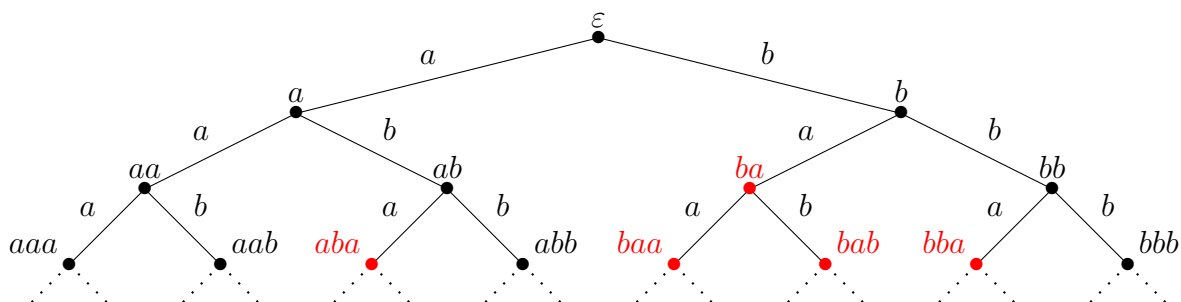


FIGURE 1.3 – Arbre lexicographique associé au langage $L = \{a^n b^m \mid n, m \in \mathbb{N}\}$.

Soit $T = (V, E)$ un arbre labellisé par des lettres d'un alphabet \mathcal{A} . Si $w \in \mathcal{A}^*$ est la label d'un chemin dans T partant de la racine, le *suffixe* de T ayant pour racine w est noté $T[w]$. De la même façon, si $h \geq 0$ le *sous-arbre* de hauteur h et de racine w est dénoté $T[w, h]$. En particulier, si $w = \varepsilon$, alors $T[\varepsilon, h]$ est le préfixe de hauteur h de T . Le *domaine* d'un suffixe $T[w]$ est l'ensemble des labels des chemins pouvant être lu depuis le noeud w et est noté $\text{dom}(T[w])$. Si $w, w' \in \mathcal{A}^*$ alors les suffixes $T[w]$ et $T[w']$ sont dit *isomorphes* si $\text{dom}(T[w]) = \text{dom}(T[w'])$. Si les noeuds sont décorés, deux sous-arbres sont isomorphes s'ils ont le même domaine et les mêmes décorations.

Exemple 1.3.7. Considérons l'arbre de la Figure 1.3. Cet arbre ne possède que 3 sous-arbres non isomorphes donnés par \mathcal{T}_L lui-même, $\mathcal{T}_L[b]$ et $\mathcal{T}_L[ba]$ représentés à la Figure 1.4.

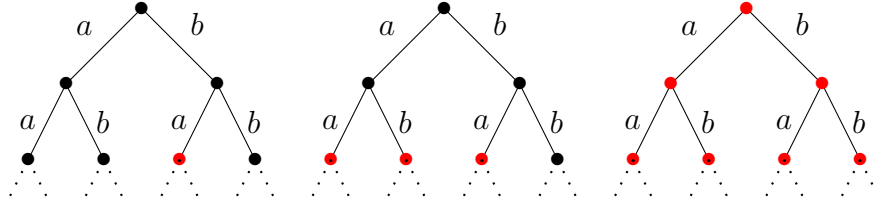


FIGURE 1.4 – Sous arbre \mathcal{T}_L , $\mathcal{T}_L[b]$ et $\mathcal{T}_L[ba]$ de l'arbre lexicographique associé à $L = \{a^n b^m \mid n, m \in \mathbb{N}\}$.

- Le sous-arbre \mathcal{T}_L correspond à la lecture de ε dans l'arbre. Les noeuds colorés en noir depuis cette racine font partie de $\varepsilon^{-1}L = L$.
- Le sous-arbre $\mathcal{T}_L[b]$ correspond à la lecture de b dans l'arbre. Les noeuds colorés en noir depuis cette racine font partie de $b^{-1}L = \{b^m \mid m \in \mathbb{N}\}$.
- Le sous-arbre $\mathcal{T}_L[ba]$ correspond à la lecture de ba dans l'arbre. Les noeuds colorés en noir depuis cette racine font partie de $(ba)^{-1}L = \emptyset$.

Il n'y a aucun autre sous-arbre dans \mathcal{T}_L car les seuls mots pouvant être lus dans \mathcal{T}_L sont d'une des formes suivantes

- Les mots de la forme $\{a^n \mid n \in \mathbb{N}\}$. Ils mènent à un suffixe de la forme de $\mathcal{T}_L[a]$.
- Les mots de la forme $\{a^n b^m \mid n, m \in \mathbb{N}, m > 0\}$. Ils mènent à un suffixe de la forme de $\mathcal{T}_L[b]$.
- Les mots de la forme $\{a^n b^m a^p \mid n, m, p \in \mathbb{N}, m, p > 0\}$. Ils mènent à un suffixe de la forme de $\mathcal{T}_L[ba]$.

L'arbre lexicographique associé au langage régulier $L = \{a^n b^m \mid n, m \in \mathbb{N}\}$ a donc un nombre fini de sous-arbres non isomorphes, ce qui n'est en réalité pas une coïncidence.

Lemme 1.3.8 ([7]). Soit $L \subseteq \mathcal{A}^*$ un langage. Pour tout $u \in \mathcal{A}^*$, le sous-arbre $\mathcal{T}_{u^{-1}L}$ est isomorphe au sous-arbre $\mathcal{T}_L[u]$.

Démonstration. Pour tout $w \in \mathcal{A}^*$, la lecture de w dans $\mathcal{T}_{u^{-1}L}$ mène à un noeud dans L si et seulement si $w \in u^{-1}L$, i.e. $uw \in L$. Cela est équivalent à dire que la lecture du mot uw dans l'arbre \mathcal{T}_L mène à un noeud correspondant à un mot de L . Pour tout $w \in \mathcal{A}^*$, la lecture du mot uw dans l'arbre \mathcal{T}_L étant équivalente à la lecture du mot w dans l'arbre $\mathcal{T}_L[u]$, nous obtenons $\mathcal{T}_L[u] = \mathcal{T}_{u^{-1}L}$. \square

Théorème 1.3.9. *Un langage L est régulier si et seulement l'arbre lexicographique associé à un nombre fini de sous-arbres non isomorphes.*

Le nombre de suffixes non isomorphes de l'arbre \mathcal{T}_L est donc égal au nombre de classes d'équivalence pour la congruence de Nérode. Chaque état de l'automate minimal \mathbb{A}_L correspondant au nombre de classes d'équivalence de l'équivalence de Nérode, l'automate minimal \mathbb{A}_L a un nombre fini d'états si et seulement si l'équivalence de Nérode a un nombre fini de classes d'équivalence. La finitude du nombre de suffixes implique donc la régularité du langage L et inversement par le Théorème 1.2.7.

1.3.2 Arbre associé à un langage clos par préfixe

Comme introduit dans la Section 1.1, un langage $L \subseteq \mathcal{A}^*$ est clos par préfixe si tout préfixe d'un mot de L est également un mot de L .

Considérons $L \subseteq \mathcal{A}^*$ un tel langage et définissons $<_{\mathcal{A}}$ un ordre sur l'alphabet \mathcal{A} . Nous pouvons grâce à cela ordonner les mots de L avec l'ordre radiciel noté \prec .

Définition 1.3.10. Soit $(\mathcal{A}, <_{\mathcal{A}})$ un alphabet totalement ordonné et soient $u, v \in \mathcal{A}^*$. Nous dirons que u est strictement plus petit que v pour l'ordre radiciel, noté $u \prec v$, si une des deux conditions suivantes est vérifiée :

- $|u| < |v|$
- Si $|u| = |v|$ et qu'il existe des mots $w, u', v' \in \mathcal{A}^*$ et des lettres telles que $u = wau'$ et $v = wbv'$ avec $a <_{\mathcal{A}} b$.

Nous pouvons donc associer à ce langage L un arbre, noté $T(L)$, dont les arêtes sont labellisées par les lettres de \mathcal{A} et dont les noeuds sont en bijection avec les mots de L . Contrairement à l'arbre lexicographique \mathcal{T}_L , seuls les mots de L peuvent être lus sur les chemins liant la racine aux noeuds de l'arbre $T(L)$. L'arbre $T(L)$ est construit de la manière suivante : Si deux mots u et ua sont dans L , alors une arête labellisée par a relie les noeuds correspondants de $T(L)$. Les fils d'un noeud sont ordonnés par l'ordre $<_{\mathcal{A}}$ des labels sur les arêtes.

Le label du chemin liant la racine de $T(L)$ à son n -ième noeud est exactement le n -ième mot de L ordonné par l'ordre radiciel.

Exemple 1.3.11. Considérons l'alphabet $\mathcal{A} = \{a, b\}^*$ tel que $a <_{\mathcal{A}} b$. Les premiers mots du langage $L = \{a^n b^m \mid n, m \in \mathbb{N}\}$ énumérés grâce à l'ordre radiciel sont donnés par $\varepsilon, a, b, aa, ab, bb, aaa, aab, abb, bbb \dots$. L'arbre $T(L)$ correspondant est donné à la Figure 1.5.

Inversement, nous pouvons considérer un arbre T dont les arêtes sont labellisées par l'alphabet $(\mathcal{A}, <_{\mathcal{A}})$ ordonné et dont les fils d'un noeud sont ordonnés grâce à l'ordre

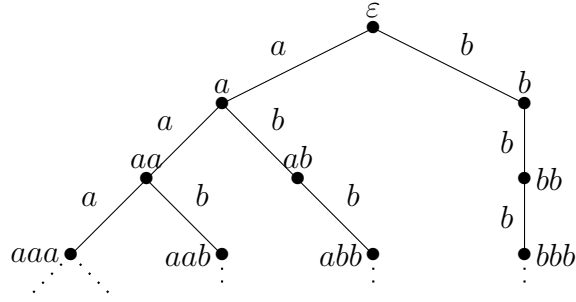


FIGURE 1.5 – Arbre $T(L)$ associé au langage $L = \{a^n b^m \mid n, m \in \mathbb{N}\}$.

$<_{\mathcal{A}}$ sur les labels des arêtes. Chaque noeud de T est alors en bijection avec le label du chemin le liant à la racine, qui est un mot de \mathcal{A}^* . L'énumération des noeuds de T (et en particulier des labels du chemin le liant à la racine) nous donne donc un langage $L \subseteq \mathcal{A}^*$ clos par préfixe. Le n -ième noeud de l'arbre T pour le parcours en largeur correspond donc au n -ième mot de L pour l'ordre radiciel sur l'alphabet ordonné $(\mathcal{A}, <_{\mathcal{A}})$.

Remarque 1.3.12. Soit $L \subseteq \mathcal{A}^*$ un langage clos par préfixe. Supposons L est *prolongeable à droite*, i.e. si $w \in L$ il existe $a \in \mathcal{A}$ tel que $wa \in L$. Dans ce cas toutes les branches de l'arbre $T(L)$ associé sont infinies.

1.3.3 Arbre associé à une signature purement périodique labellisée

À la place de construire un arbre $T(L)$ à partir d'un langage $L \subseteq \mathcal{A}^*$, nous pouvons définir un arbre à partir d'une *signature labellisée purement périodique*.

Définition 1.3.13. Soit T un arbre ordonné et \mathcal{A} un alphabet fini. Une *signature labellisée* est une suite infinie $\mathbf{s} = (w_n)_{n \geq 0}$ de mots finis non vides sur \mathcal{A} donnant une *signature* $(|w_n|)_{n \geq 0}$ et une *labellisation* de l'arbre donnée par les w_n . Une *signature* de l'arbre T est la suite des degrés des noeuds visités en parcourant l'arbre en largeur. Le n -ième noeud de l'arbre ayant un degré égal à $|w_n|$, il possède $|w_n|$ fils. Une *labellisation* est la suite des labels des arêtes de T en parcourant l'arbre en largeur. Le label de la i -ième arête du n -ième noeud de l'arbre est donnée par $w_{n,i}$ pour tout $i \in \{0, \dots, |w_n| - 1\}$.

Exemple 1.3.14. Par exemple, la signature \mathbf{s} associé à l'arbre représenté à la Figure 1.5 est la suite suivante

$$\mathbf{s} = ab, ab, b, ab, b, b, \dots, \overbrace{ab}^{n\text{-ième occurrence du terme } ab \text{ dans la suite } \mathbf{s}}, \underbrace{b \dots b}_{n-1}, \dots$$

En particulier, la signature labellisée \mathbf{s} peut être purement périodique et est dans ce cas appelée *signature labellisée purement périodique*. L'arbre associé à la signature labellisée \mathbf{s} est noté $T(\mathbf{s})$. Le langage clos par préfixe associé à l'arbre $T(\mathbf{s})$ est noté $L(\mathbf{s})$.

Exemple 1.3.15. L'arbre représenté à la Figure 1.3 non décoré est l'arbre $T(\mathbf{s})$ associé à la signature purement périodique $\mathbf{s} = (ab)^\omega$.

Un *i-arbre* est un arbre particulier dont la racine est un fils d'elle-même. Bien que les boucles soient normalement interdites dans la définition d'un arbre, l'ajout d'une boucle labellisée par 0 se révèle régulièrement utile lorsqu'on travaille avec des systèmes de numération pour traiter les 0 de tête éventuels présents dans la représentation des nombres naturels. Nous pouvons également associer à une signature labellisée purement périodique un arbre de ce type.

Exemple 1.3.16. Considérons la signature périodique labellisée $\mathbf{s} = (02, 1)^\omega$ et notons $T(\mathbf{s})$ l'*i*-arbre associé à la signature labellisée purement périodique \mathbf{s} . Au vu de la définition d'une signature périodique labellisée, les noeuds de l'*i*-arbre $T(\mathbf{s})$ ont successivement 2 fils puis 1 fils. Plus précisément, pour tout $n \geq 0$

$$\text{le noeud } n \text{ a } \begin{cases} 2 \text{ fils} & \text{si } n \equiv 0 \pmod{2}, \\ 1 \text{ fils} & \text{si } n \equiv 1 \pmod{2}. \end{cases}$$

De plus, la première arête sortante des noeuds de degré 2 est labellisée par 0. L'autre est labellisée par 2. L'unique arête sortante des noeuds de degré 1 est labellisée par 1. L'arbre associé $T(\mathbf{s})$ étant un *i*-arbre, sa racine est un fils d'elle-même, i.e. il y a une boucle sur la racine. Cette boucle est labellisée par 0 car $0 \equiv 0 \pmod{2}$. Une représentation des premiers niveaux de $T(\mathbf{s})$ est donnée à la Figure 1.6.

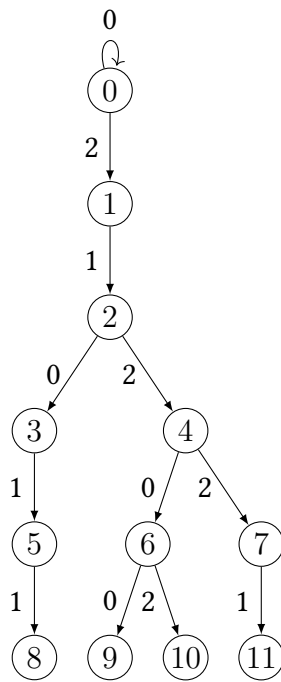


FIGURE 1.6 – Arbre associé à la signature périodique labellisée $\mathbf{s} = (02, 1)^\omega$.

Chapitre 2

Système de numération

Dans ce chapitre, nous définissons précisément plusieurs systèmes de numération, à commencer par les systèmes de numération abstraits. Ensuite, nous introduisons les systèmes de numération en bases rationnelles, qui en sont un cas particulier. Nous abordons également les systèmes de numération positionnels, et faisons des liens entre tous ces systèmes de numération.

2.1 Systèmes de numération abstraits

Considérons un langage $L \subseteq \mathcal{A}^*$ infini sur un alphabet $(\mathcal{A}, <_{\mathcal{A}})$ totalement ordonné. L'alphabet \mathcal{A} étant totalement ordonné, cela permet d'énumérer les mots de L dans l'ordre radiciel et d'y associer un système de numération. Les systèmes de numération ainsi construits sont appelés *systèmes de numération abstraits*.

Définition 2.1.1. Un *système de numération abstrait*, noté SNA, est un triplet $\mathcal{S} = (L, \mathcal{A}, <_{\mathcal{A}})$ où L est un langage infini sur l'alphabet $(\mathcal{A}, <_{\mathcal{A}})$ totalement ordonné. Le langage L est appelé *langage de numération*.

L'application $\text{rep}_{\mathcal{S}} : \mathbb{N} \rightarrow L$ est la bijection qui à $n \in \mathbb{N}$ associe le $(n + 1)$ -ième mot de L , ordonné par l'ordre radiciel. Ce mot est appelé la *\mathcal{S} -représentation* de n . La \mathcal{S} -représentation de 0 est le premier mot de L .

L'application inverse est $\text{val}_{\mathcal{S}} : L \rightarrow \mathbb{N}$. Pour tout $w \in L$, $\text{val}_{\mathcal{S}}(w)$ s'appelle la *valeur* dans le système de numération abstrait \mathcal{S} .

Comme introduit dans la Section 1.3.2, un langage clos par préfixe L peut être associé à un arbre dont les arêtes sont labellisées par des lettres d'un alphabet totalement ordonné $(\mathcal{A}, <_{\mathcal{A}})$. Les arêtes étant ordonnées par l'ordre $<_{\mathcal{A}}$ sur les labels, l'énumération des noeuds de l'arbre grâce au parcours en largeur donne une énumération des mots de L dans l'ordre radiciel. Cette énumération des mots de L nous permet en particulier de définir un système de numération abstrait.

n	$\text{rep}_{\mathcal{S}}(n)$
0	ε
1	2
2	21
3	210
4	212
5	2101
6	2120

n	$\text{rep}_{\mathcal{S}}(n)$
7	2122
8	21011
9	21200
10	21202
11	21221
12	210110
13	210112

n	$\text{rep}_{\mathcal{S}}(n)$
14	212001
15	212020
16	212022
17	212211
18	2101100
19	2101102
20	2101121

TABLE 2.1 – Représentation des premiers naturels dans le système de numération \mathcal{S} .

Exemple 2.1.2. Considérons l'arbre représenté à la Figure 1.6 dans lequel est omis la boucle sur la racine. Les premiers mots du langage L associé sont les mots suivants

$$\varepsilon, 2, 21, 210, 212, 2101, 2120, \dots$$

Nous pouvons définir un système de numération abstrait $\mathcal{S} = (L, \mathcal{A}_3, <)$ où $<$ est l'ordre sur les nombres naturels. La représentation des premiers naturels dans le système de numération abstrait \mathcal{S} sont donnés dans la Table 2.1.

2.2 Base rationnelle

Dans cette section, nous introduisons le concept de base rationnelle, apparaissant pour la première fois dans [1]. Nous considérons des entiers $p > q > 1$ premiers entre eux, de façon à ce que $\frac{p}{q}$ soit irréductible et supérieure à 1. La fraction $\frac{p}{q}$ est appelée la *base* du système de numération.

2.2.1 Fonction d'évaluation

Définition 2.2.1. Soient p et q tels que $p > q > 1$ des entiers et $n \in \mathbb{N}$. La valeur d'un mot $w \in \mathcal{A}_p^*$ est donnée par la *fonction d'évaluation en base rationnelle* définie comme :

$$\text{val}_{\frac{p}{q}}(\cdot) : \mathcal{A}_p^* \rightarrow \mathbb{Q}$$

$$w = w_m \cdots w_0 \mapsto \text{val}_{\frac{p}{q}}(w) = \sum_{i=0}^m \frac{w_i}{q} \left(\frac{p}{q}\right)^i$$

où les chiffres de poids les plus faibles sont à droite.

Cette définition n'implique donc pas forcément que l'évaluation en base rationnelle d'un mot de \mathcal{A}_p^* est un entier. Par exemple, $\text{rep}_{\frac{3}{2}}(22) = \frac{3}{2} + 1 = 2.5$.

Avec la fonction d'évaluation en base rationnelle, nous pouvons déduire quelques propriétés. Ces formules découlent directement de la définition de la fonction $\text{val}_{\frac{p}{q}}(\cdot)$. Pour tout $w, v \in \mathcal{A}_p^*$ et pour tous $p > q > 1$ entiers,

$$\text{val}_{\frac{p}{q}}(aw) = \frac{a}{q} \cdot \left(\frac{p}{q}\right)^{|w|} + \text{val}_{\frac{p}{q}}(w) \quad (2.1)$$

$$\text{val}_{\frac{p}{q}}(wa) = \text{val}_{\frac{p}{q}}(w) \cdot \frac{p}{q} + \frac{a}{q} \quad (2.2)$$

$$\text{val}_{\frac{p}{q}}(wv) = \text{val}_{\frac{p}{q}}(w) \cdot \left(\frac{p}{q}\right)^{|v|} + \text{val}_{\frac{p}{q}}(v) \quad (2.3)$$

2.2.2 Représentation en base $\frac{p}{q}$

À tout entier $n \in \mathbb{N}$ peut être associé une *représentation en base $\frac{p}{q}$* . Celle-ci est notée $\text{rep}_{\frac{p}{q}}(n)$ et s'obtient grâce à l'*algorithme d'Euclide généralisé* décrit ci dessous :

Posons $N_0 = n$. Ensuite, nous pouvons définir la suite $(N_i)_{i \geq 0}$ comme

$$\begin{aligned} qN_0 &= pN_1 + a_0 \\ qN_1 &= pN_2 + a_1 \\ &\vdots \\ qN_{m-1} &= pN_m + a_{m-1} \\ qN_m &= p \cdot 0 + a_m \end{aligned}$$

Pour tout $i \geq 1$, a_i est le reste de la division euclidienne de qN_{i-1} par p ce qui implique $a_i \in \mathcal{A}_p$ pour tout $i \geq 1$. De plus, N_{i+1} étant obtenu par division euclidienne de qN_i par $p > q$, la suite $(N_i)_{i \geq 0}$ est strictement décroissante et il existe $m \in \mathbb{N}$ pour lequel la suite devient stationnaire en la valeur 0, i.e. il existe $m \in \mathbb{N}$ tel que $N_{m+1} = N_{m+2} = \dots = 0$.

Définition 2.2.2. Soit $n \in \mathbb{N}$ et des entiers $p > q > 1$, p, q premiers entre eux. La représentation en base $\frac{p}{q}$ de n est le mot sur \mathcal{A}_p déterminé grâce à l'algorithme d'Euclide généralisé. En utilisant les notations introduites précédemment, nous avons donc

$$\text{rep}_{\frac{p}{q}}(n) = a_m \cdots a_0,$$

avec $a_0, \dots, a_m \in \mathcal{A}_p$, $a_m \neq 0$.

Exemple 2.2.3. La représentation en base $\frac{3}{2}$ de 10 se calcule de la manière suivante :
 $N_0 = 10$

$$\begin{aligned} 2N_0 &= 3 \cdot 6 + 2 \Rightarrow N_1 = 6 \text{ et } a_0 = 2 \\ 2N_1 &= 3 \cdot 4 + 0 \Rightarrow N_2 = 4 \text{ et } a_1 = 0 \\ 2N_2 &= 3 \cdot 2 + 2 \Rightarrow N_3 = 2 \text{ et } a_2 = 2 \\ 2N_3 &= 3 \cdot 1 + 1 \Rightarrow N_4 = 1 \text{ et } a_3 = 1 \\ 2N_4 &= 3 \cdot 0 + 2 \Rightarrow N_5 = 0 \text{ et } a_4 = 2 \end{aligned}$$

Nous avons donc finalement,

$$\text{rep}_{\frac{3}{2}}(10) = 20212$$

Remarque 2.2.4. Pour tous $p > q > 1$, la représentation de 0 en base $\frac{p}{q}$ est ε . En effet, le terme pour lequel la suite $(N_i)_{i \geq 0}$ devient stationnaire en 0 est dans ce cas N_0 . L'algorithme d'Euclide généralisé s'arrête avant de définir les lettres a_i formant la représentation de 0. Sa représentation en base $\frac{p}{q}$ est donc le mot vide.

Remarque 2.2.5. Remarquons que dans le cas où $q = 1$ (interdit dans la définition d'une base rationnelle), l'algorithme d'Euclide généralisé coïncide avec l'algorithme d'Euclide classique permettant d'obtenir la décomposition en base entière p . Les fonctions val_p et $\text{val}_{\frac{p}{q}}$ sont elles aussi identiques.

Les bases rationnelles et entières partagent donc certaines propriétés, dont la suivante.

Proposition 2.2.6. Soit $n \in \mathbb{N}$ tel que $\text{rep}_{\frac{p}{q}}(n) = a_m \cdots a_0$. Alors nous avons la propriété suivante

$$\text{val}_{\frac{p}{q}}(\text{rep}_{\frac{p}{q}}(n)) = n.$$

Démonstration. Etant donné que $\text{rep}_{\frac{p}{q}}(n) = a_m \cdots a_0$, nous savons que

$$\text{val}_{\frac{p}{q}}(\text{rep}_{\frac{p}{q}}(n)) = \sum_{i=0}^m \frac{a_i}{q} \left(\frac{p}{q}\right)^i.$$

Nous savons également que la représentation en base rationnelle de n , $\text{rep}_{\frac{p}{q}}(n)$, est obtenue grâce à l'algorithme d'Euclide généralisé, c'est-à-dire,

$$\begin{aligned} n &= N_0 \\ qN_0 &= pN_1 + a_0 \\ &\vdots \\ qN_m &= pN_{m+1} + a_m. \end{aligned}$$

De ce fait, nous avons,

$$\begin{aligned}
n = N_0 &= \frac{p}{q}N_1 + \frac{a_0}{q} \\
&= \frac{p}{q}\left(\frac{p}{q}N_2 + a_1\right) + \frac{a_0}{q} \\
&= \left(\frac{p}{q}\right)^2 N_2 + \frac{p a_1}{q} + \frac{a_0}{q} \\
&\vdots \\
&= \left(\frac{p}{q}\right)^i N_i + \left(\frac{p}{q}\right)^{i-1} \frac{a_{i-1}}{q} + \dots + \frac{a_0}{q} \\
&\vdots \\
&= \left(\frac{p}{q}\right)^m N_m + \left(\frac{p}{q}\right)^{m-1} \frac{a_{m-1}}{q} + \dots + \frac{a_0}{q} \\
&= \sum_{i=0}^m \frac{a_i}{q} \left(\frac{p}{q}\right)^i.
\end{aligned}$$

car $N_{m+1} = 0$ nous donne $N_m = \frac{a_m}{q}$. Cela conclut la démonstration au vu de l'égalité des deux membres. □

Nous avons également la Proposition suivante, affirmant que deux mots ayant la même valeur en base $\frac{p}{q}$ sont identiques à des zéros de tête près.

Proposition 2.2.7. *Soient deux mots u, v sur l'alphabet \mathcal{A}_p . Si $\text{val}_{\frac{p}{q}}(u) = \text{val}_{\frac{p}{q}}(v)$, alors il existe i tel que $u = 0^i v$ ou $v = 0^i u$.*

Nous pouvons associer à tout $p > q > 1$ entiers premiers entre eux le langage $L_{\frac{p}{q}} = \{\text{rep}_{\frac{p}{q}}(n) \mid n \in \mathbb{N}\}$ constitué des représentations en base $\frac{p}{q}$ des naturels.

Remarque 2.2.8. Contrairement au langage $L_p = \{\text{rep}_p(n) \mid n \in \mathbb{N}\}$, le langage $L_{\frac{p}{q}}$ n'est pas régulier. En effet, comme démontré dans [16, Théorème 4.10], le langage $L_{\frac{p}{q}}$ n'est même pas algébrique.

Néanmoins, ce langage a certaines propriétés intéressantes que nous allons quelque peu détailler.

Proposition 2.2.9. *Le langage $L_{\frac{p}{q}}$ est clos par préfixe et prolongeable à droite.*

Démonstration. Soit $u \in L_{\frac{p}{q}}$ tel que $u \neq \varepsilon$ et soit $n \in \mathbb{N}$ l'entier qui vérifie $\text{rep}_{\frac{p}{q}}(n) = u$. L'entier n est donc en particulier différent de 0.

Soient $a \in \mathcal{A}_p$ et n' l'entier qui vérifie $qn = pn' + a$. Nous savons donc que

$$\text{rep}_{\frac{p}{q}}(n) = \text{rep}_{\frac{p}{q}}(n')a,$$

ce qui implique que le langage $L_{\frac{p}{q}}$ est clos par préfixe.

Montrons maintenant que le langage $L_{\frac{p}{q}}$ est prolongeable à droite. Au vu du fait que

$p > q$, il existe au moins un $a \in \mathcal{A}_p$ qui vérifie $pn + a \equiv 0 \pmod{q}$. Dans ce cas $m = \frac{np+a}{q}$ est un entier supérieur à n et non nul, qui vérifie $qm = pn + a$. Nous avons donc également

$$\text{rep}_{\frac{p}{q}}(m) = \text{rep}_{\frac{p}{q}}(n)a. \quad (2.4)$$

L'égalité 2.4 découle de l'algorithme d'Euclide généralisé. En effet, si

$$qm = pn + a$$

alors la décomposition de m en base $\frac{p}{q}$ est telle que $a_0 = a$. La suite de la décomposition de m en base $\frac{p}{q}$ est alors donnée par la décomposition de n en base $\frac{p}{q}$ car $n = N_1$ dans l'algorithme d'Euclide généralisé. L'égalité (2.4) implique que $L_{\frac{p}{q}}$ est prolongeable à droite, ce qui conclut la preuve. \square

La Proposition 2.2.9 nous apprend donc qu'un arbre peut être associé au langage $L_{\frac{p}{q}}$, car c'est un langage clos par préfixe. De plus, chacune des branches de $T(L_{\frac{p}{q}})$ est infinie car le langage est prolongeable à droite.

Proposition 2.2.10. *Soient deux entiers m et n . Nous avons que $n < m$ si et seulement si $\text{rep}_{\frac{p}{q}}(n) \prec \text{rep}_{\frac{p}{q}}(m)$.*

Démonstration. Notons $\text{rep}_{\frac{p}{q}}(n) = a_{k-1} \cdots a_0$ et $\text{rep}_{\frac{p}{q}}(m) = b_i \cdots b_0$.

Condition suffisante : Si nous supposons que $\text{rep}_{\frac{p}{q}}(n) \prec \text{rep}_{\frac{p}{q}}(m)$, nous savons que $k - 1 \leq i$. Montrons par récurrence sur k que $n < m$. Si $k = 0$ alors $\text{rep}_{\frac{p}{q}}(n) = \varepsilon$ et donc $n = 0$. Puisque $\text{rep}_{\frac{p}{q}}(n) \prec \text{rep}_{\frac{p}{q}}(m)$, $\text{rep}_{\frac{p}{q}}(m)$ n'est pas le mot vide et donc

$$m = \text{val}_{\frac{p}{q}}(\text{rep}_{\frac{p}{q}}(m)) > 0 = n.$$

Supposons maintenant que $1 \leq k - 1 \leq i$. Puisque $L_{\frac{p}{q}}$ est clos par préfixe, $n' = \text{val}_{\frac{p}{q}}(a_{k-1} \cdots a_1)$ et $m' = \text{val}_{\frac{p}{q}}(b_i \cdots b_1)$ sont des entiers. En utilisant la propriété 2.2, nous obtenons,

$$n - m = (n' - m')\frac{p}{q} + \frac{a_0 - b_0}{q}. \quad (2.5)$$

Deux cas sont maintenant possibles puisque $\text{rep}_{\frac{p}{q}}(n) \prec \text{rep}_{\frac{p}{q}}(m)$:

- Si $\text{rep}_{\frac{p}{q}}(n') \prec \text{rep}_{\frac{p}{q}}(m')$, alors par hypothèse de récurrence $n' - m' < 0$. Cela implique que $m' - n' > 0$ et donc $m' - n' \geq 1$. De plus, peu importe les lettres a_0 et b_0 nous avons

$$-(p - 1) \leq a_0 - b_0 \leq p - 1,$$

ce qui implique

$$\frac{a_0 - b_0}{q} < \frac{p}{q} \leq \frac{p}{q}(m' - n').$$

En utilisant l'égalité (2.5), nous obtenons donc

$$n - m = (n' - m')\frac{p}{q} + \frac{a_0 - b_0}{q} < (n' - m')\frac{p}{q} + (m' - n')\frac{p}{q} = 0,$$

et donc $n - m < 0$ comme souhaité.

- Si $\text{rep}_{\frac{p}{q}}(n') = \text{rep}_{\frac{p}{q}}(m')$ et $a_0 < b_0$, alors

$$\text{val}_{\frac{p}{q}}(\text{rep}_{\frac{p}{q}}(n')) = \text{val}_{\frac{p}{q}}(\text{rep}_{\frac{p}{q}}(m')) \Rightarrow n' = m',$$

et en utilisant l'égalité (2.5)

$$n - m = \frac{1}{q}(a_0 - b_0) < 0.$$

□

De la Proposition 2.2.10 découle que pour tout $n \in \mathbb{N}$, la représentation de n en base $\frac{p}{q}$ est le $(n + 1)$ -ième mot du langage $L_{\frac{p}{q}}$. De cette façon, un système de numération abstrait $\mathcal{S} = (L_{\frac{p}{q}}, \mathcal{A}_p, <)$ peut être défini grâce aux langages $L_{\frac{p}{q}}$. Les systèmes de numération en bases rationnelles sont donc des cas particuliers de SNA sur le langage $L_{\frac{p}{q}}$.

Remarque 2.2.11. Le langage L de l'Exemple 2.1.2 n'est autre que $L_{\frac{3}{2}}$. En réalité, l'arbre $T(L_{\frac{p}{q}})$ associé au langage $L_{\frac{p}{q}}$ a une signature labellisée qui est purement périodique pour tous $p > q > 1$ premiers entre eux.

2.3 Systèmes de numération positionnels

Les systèmes de numération à bases rationnelles et à bases entières introduits dans la Section 2.2 sont tous deux des cas particuliers de *systèmes de numération positionnels*, notés SNP.

Les SNP sont basés sur une suite $U = (U_n)_{n \geq 0}$ strictement croissante d'entier, telle que $U_0 = 1$ et telle que la suite $\left(\frac{U_n}{U_{n+1}}\right)_{n \geq 0}$ est bornée par une constante C . Tout entier $n \in \mathbb{N}$ peut être représenté grâce à cette suite à l'aide d'un algorithme glouton qui est une alternative à celui utilisé dans la Section 2.2.2. Soit $n \in \mathbb{N}$ un entier naturel, pour lequel il existe un unique $m \in \mathbb{N}$ tel que $U_m \leq n < U_{m+1}$. Notons $n := r_{m+1}$. En effectuant la division euclidienne de n par U_m et en itérant sur le reste de cette division, nous pouvons définir la suite de reste $(r_i)_{i \geq 0}$ de la façon suivante

$$\begin{aligned}
n &= a_m U_m + r_m \quad \text{avec } r_m < U_m \\
r_m &= a_{m-1} U_{m-1} + r_{m-1} \\
&\vdots \\
r_2 &= a_1 U_1 + r_1 \\
r_1 &= a_0 U_0 + \underbrace{r_0}_{\text{Vaut toujours 0 car } U_0 = 1.}
\end{aligned}$$

L'algorithme précédent nous permet donc de décomposer le nombre n comme suit

$$n = \sum_{i=0}^m a_i U_i.$$

Le mot $a_m \cdots a_0$ est appelé *U -représentation normale* de n et noté $\text{rep}_U(n)$. Comme dans le cas de la représentation en base rationnelle, le nombre 0 est représenté par ε et nous notons $L_U = \{\text{rep}_U(n) \mid n \in \mathbb{N}\}$.

La *U -valeur* d'un mot $w = w_m \cdots w_0$ sur \mathbb{Z} , notée $\text{val}_U(w)$, est définie comme

$$\begin{aligned}
\text{val}_U(\cdot) : \mathbb{N}^* &\rightarrow \mathbb{N} \\
w = w_m \cdots w_0 &\mapsto \text{val}_U(w) = \sum_{i=0}^m w_i U_i
\end{aligned}$$

Si la U -valeur d'un mot w est égale à un certain entier n , alors w est une U -représentation de n . Cette représentation n'est pas nécessairement la même que la U -représentation normale de cet entier n .

Remarque 2.3.1. • La condition $U_0 = 1$ sur le premier terme de la suite U permet que tout entier n admette au moins une U -représentation normale grâce à la suite U donnée par $n = nU_0$.

- Le fait que la suite $\left(\frac{U_n}{U_{n+1}}\right)_{n \geq 0}$ soit bornée par une constante C permet que l'alphabet du langage L_U soit fini. Plus précisément, nous appelons $\mathcal{A}_U = \{0, \dots, b\}$ l'alphabet minimal tel que $L_U \subseteq \mathcal{A}_U^*$ et tel que b est le plus grand entier plus petit que $\sup_{n \geq 0} \frac{U_n}{U_{n+1}}$.
- Le système de numération en base entière k est un cas particulier de système de numération positionnel avec $U = (k^n)_{n \geq 0}$.

Considérons un SNP $U = (U_n)_{n \geq 0}$. La suite U peut être définie de façon à vérifier certaines conditions, comme respecter une relation de récurrence linéaire à coefficients constants. Dans ce cas, le système de numération positionnel qui lui est associé est alors appelé *système de numération positionnel linéaire*. Un système de numération est appelé

un système de numération de Bertrand si pour tout $w \in \mathcal{A}_U^+$, $w \in L_U$ si et seulement si $w0 \in L_U$.

Les systèmes de numération positionnels vérifient également la Proposition 2.2.10 démontrée dans le cas des systèmes de numération à bases rationnelles. Les systèmes de numération positionnels sont donc également des cas particuliers de systèmes de numération abstraits. En effet, pour définir un SNA il nous suffit d'avoir un triplet $(L, \mathcal{A}, <_{\mathcal{A}})$ afin de que les mots de L puissent être ordonnés avec l'ordre radiciel. Pour les systèmes de numération positionnels U , il suffit donc de prendre $L = L_U$, $\mathcal{A} = \mathcal{A}_U$. Les lettres de \mathcal{A}_U sont ordonnées par l'ordre $<$ sur les nombres naturels, il suffit donc de choisir $<_{\mathcal{A}} = <$.

Exemple 2.3.2. Le système de numération de Fibonacci, aussi appelé système de numération de Zeckendorf, illustre le concept de système de numération positionnel linéaire. La suite U est définie de la façon suivante

$$\begin{cases} U_0 = 1 \\ U_1 = 2 \\ U_{n+2} = U_{n+1} + U_n \quad \text{pour tout } n \geq 0, \end{cases}$$

Sachant que la suite $\left(\frac{U_n}{U_{n+1}}\right)_{n \geq 0}$ converge vers le nombre d'or φ , l'alphabet $A_U = \{0, 1\}$. Le nombre 11110 est une U -représentation du nombre 18 dans le système de numération de Fibonacci et la U -représentation normale de 18 est donnée par 101000. Le langage $L_U = 1\{0, 01\}^*$ car les nombres naturels sont représentés par l'algorithme glouton. Ce système de numération est un exemple de système de numération de Bertrand car si $w \in L_U$ est un mot non vide, alors $w0 \in L_U$ au vu de la forme du langage L_U .

Nous pouvons cependant remarquer que bien que tous les systèmes de numérations positionnels soient des cas particuliers de systèmes de numération abstraits, l'inverse n'est en général pas vrai.

Exemple 2.3.3 ([21], Remarque 2.43). Considérons le SNA $\mathcal{S} = (a^*b^*, \{a, b\}, a < b)$. Les premiers mots du langage $L = a^*b^*$ dans l'ordre radiciel sont donnés par

$$\varepsilon \prec a \prec b \prec aa \prec ab \prec bb \dots$$

Supposons que nous avons une application $v : \{a, b\} \rightarrow \mathbb{N}$ attribuant un poids aux lettres a, b . Nous allons montrer qu'il n'existe aucune suite $(U_n)_{n \geq 0}$ définissant un système de numération positionnel tel que pour tout mot $a_k \cdots a_0 \in \{a, b\}^*$

$$\text{val}_{\mathcal{S}}(a_k \cdots a_0) = \text{val}_U(v(a_k) \cdots v(a_0)) := \sum_{i=0}^k v(a_i)U_i. \quad (2.6)$$

Supposons qu'il existe une suite U vérifiant la condition 2.6. Dans la système de numération abstrait \mathcal{S} , nous avons $\text{rep}_{\mathcal{S}}(1) = a$ et $\text{rep}_{\mathcal{S}}(2) = b$. En ajoutant à cela la condition $U_0 = 1$, cela implique $\text{val}_{\mathcal{S}}(a) = v(a)U_0$ et donc $v(a) = 1$. De la même façon, nous avons $v(b) = 2$. Maintenant que les poids attribués aux lettres a, b sont fixés, nous pouvons déduire la valeur du terme U_1 de la suite U . En effet, nous devons avoir l'égalité

$$\text{val}_{\mathcal{S}}(aa) = 3 = v(a)U_1 + v(a)U_0,$$

ce qui implique $U_1 = 2$. De ce fait,

$$\begin{aligned} \text{val}_U(v(b)v(b)) &= v(b)U_1 + v(b)U_0 \\ &= 2.2 + 2.1 = 6 \end{aligned}$$

Ceci contredit l'égalité (2.6) car $\text{val}_{\mathcal{S}}(bb) = 5$.

Chapitre 3

Suites $\frac{p}{q}$ -automatiques

Dans ce Chapitre nous allons commencer par généraliser la notion de suite k -automatique à un système de numération abstrait \mathcal{S} construit sur un langage régulier afin de définir les suites \mathcal{S} -automatiques. Cette nouvelle définition mène naturellement à une généralisation du théorème de Cobham de 1972 concernant les suites k -automatiques aux suites \mathcal{S} -automatique. De là, nous introduisons une autre généralisation de ce théorème aux suites \mathcal{S} -automatiques où \mathcal{S} est un SNA dont le langage est associé à un arbre ayant une signature labellisée purement périodique. Les suites $\frac{p}{q}$ -automatiques étant un cas particulier de SNA pour $\mathcal{S} = (L_{\frac{p}{q}}, \mathcal{A}_p, <)$, ce théorème généralise le théorème de Cobham aux suites $\frac{p}{q}$ -automatiques. Dans la dernière section, nous caractérisons la $\frac{p}{q}$ -automaticité d'une suite \mathbf{x} en faisant usage du nombre de sous-arbres de $T_{\mathbf{x}}(L_{\frac{p}{q}})$.

3.1 Généralisations d'un théorème de Cobham

Considérons \mathcal{A}, \mathcal{B} des alphabets finis et soit $k \geq 2$ un entier. Un morphisme $\sigma : \mathcal{A}^* \rightarrow \mathcal{B}^*$ est une application vérifiant $\sigma(ab) = \sigma(a)\sigma(b)$ pour tout $a, b \in \mathcal{A}^1$. Il est dit k -uniforme si pour tout $a \in \mathcal{A}$, nous avons $|\sigma(a)| = k$. En particulier, un *codage* est un morphisme 1-uniforme. Une façon classique de construire des mots infinis \mathbf{x} est d'utiliser ces morphismes, et plus précisément les morphismes *prolongeables*. Un morphisme $\sigma : \mathcal{A}^* \rightarrow \mathcal{A}^*$ est prolongeable en $a \in \mathcal{A}$ si $\sigma(a) = au$ où $u \in \mathcal{A}^+$ et si $\lim_{n \rightarrow +\infty} |\sigma^n(a)| = +\infty$. On dit que σ est *effaçant* s'il existe $a \in \mathcal{A}$ tel que $\sigma(a) = \varepsilon$. Si le morphisme $\sigma : \mathcal{A}^* \rightarrow \mathcal{A}^*$ est prolongeable en $a \in \mathcal{A}$, la suite $(\sigma^n(a))_{n \geq 0}$ converge dans $\mathcal{A}^{\mathbb{N}}$ et la limite de la suite est un mot infini noté $\sigma^\omega(a)$. La limite $\sigma^\omega(a)$ est un *point fixe* du morphisme σ , i.e. $\sigma(\sigma^\omega(a)) = \sigma^\omega(a)$. De façon explicite, le point fixe du

1. Il suffit donc de connaître l'image de chacune des lettres par σ pour définir le morphisme sur tout \mathcal{A}^* .

morphisme σ commençant par a s'écrit de la façon suivante

$$\sigma^\omega(a) := au\sigma(u)\sigma^2(u)\sigma^3(u)\cdots\sigma^n(u)\cdots. \quad (3.1)$$

Exemple 3.1.1. Considérons le morphisme σ 2-uniforme défini de la façon suivante

$$\sigma : \begin{cases} 0 & \mapsto 01 \\ 1 & \mapsto 10. \end{cases}$$

Ce morphisme est prolongeable en 0. La limite $\sigma^\omega(0)$ existe donc et n'est autre que le mot de Thue–Morse $\mathbf{t} = 01101001\cdots$. Le mot infini \mathbf{t} est donc un point fixe du morphisme 2-uniforme σ .

Soient $\mathbf{x} \in \mathcal{B}^{\mathbb{N}}$ un mot infini et U un système de numération. Le mot \mathbf{x} est dit *U-automatique* s'il existe un AFDS $\mathbb{A} = (Q, q_0, \tau, \mathcal{A}_U, \delta, \mathcal{B})$ tel que $x_n = \tau(\delta(q_0, \text{rep}_U(n)))$ pour tout $n \geq 0$.

Soit $k \geq 2$ un entier. La notion de *k-automatisme* est en réalité un cas particulier de *U-automatisme* où U est le système de numération positionnelle associé à la suite $U = (k^n)_{n \geq 0}$ (Voir Section 2.3 pour plus de détails). Pour tout $n \geq 0$ entier, nous avons dans ce cas $\text{rep}_U(n) = \text{rep}_k(n)$. Dans l'article [10] de 1972, Alan Cobham démontre un théorème liant les mots *k-automatiques* et les mots qui sont points fixes par un codage d'un morphisme *k-uniforme*. Plus précisément, le théorème de Cobham de 1972 s'énonce comme suit.

Théorème 3.1.2 (Cobham,[10]). *Soient $k \geq 2$ et \mathcal{A}, \mathcal{B} deux alphabets finis. Alors le mot $\mathbf{x} \in \mathcal{B}^{\mathbb{N}}$ est *k-automatique* si et seulement si c'est l'image par un codage d'un point fixe d'un morphisme *k-uniforme*, i.e. il existe un morphisme *k-uniforme* $\sigma : \mathcal{A}^* \rightarrow \mathcal{A}^*$ et $a \in \mathcal{A}$ tels que σ est prolongeable en a et il existe un codage $\tau : \mathcal{A} \rightarrow \mathcal{B}$ tels que $\mathbf{x} = \tau(\sigma^\omega(a))$.*

Exemple 3.1.3. Considérons le mot de Thue-morse \mathbf{t} . Le Théorème 3.1.2 associe à ce mot l'AFDS représenté à la Figure 3.1. Pour passer de l'état 0 à l'état 1 de l'automate, il faut lire le chiffre 1 lors de la lecture d'un mot $w \in \mathcal{A}_2^*$. S'il y a un nombre impair (resp. pair) de 1 dans w , sa lecture s'achève dans l'état 1 (resp. 0) et l'AFDS sort le chiffre 1 (resp.0).

Au vu de la forme de l'AFDS engendrant le mot de Thue-Morse, celui-ci peut également être défini pour tout $n \geq 0$ comme $t_n = |\text{rep}_2(n)|_1 \bmod 2$.

3.1.1 Cas de systèmes de numération abstraits

La généralisation de la *k-automatisme* d'un mot infini à un système de numération quelconque a mené à de nombreuses généralisations du théorème de Cobham 3.1.2.

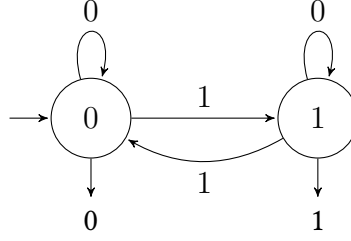


FIGURE 3.1 – L’AFDS engendrant le mot de Thue-Morse.

Le cas où le système de numération positionnelle U est un système de numération de Bertrand a, par exemple, été étudié dans [6]. Nous allons explorer dans cette section la généralisation de ce théorème lorsque le système de numération est un système de numération abstrait construit sur un langage régulier.

Dans le cas de cette généralisation, nous parlons de \mathcal{S} -automaticité où $\mathcal{S} = (L, \mathcal{A}, <_{\mathcal{A}})$ est un SNA construit sur un langage L régulier. Plus précisément, le mot $\mathbf{x} \in \mathcal{B}^{\mathbb{N}}$ est \mathcal{S} -automatic si il existe un AFDS $\mathbb{A} = (Q, q_0, \tau, \mathcal{A}, \delta, \mathcal{B})$ tel que pour tout $n \geq 0$ nous avons $x_n = \tau(\delta(q_0, \text{rep}_{\mathcal{S}}(n)))$.

Les systèmes de numération abstraits étant une généralisation des systèmes de numération en bases entières, la généralisation du théorème de Cobham aux SNA construits sur un langage régulier ne fait plus le lien entre l’automaticité et le fait d’être l’image par un codage d’un point fixe d’un morphisme uniforme. Nous allons relâcher les conditions sur les morphismes σ et τ du Théorème 3.1.2 :

- Le codage τ devient un *codage faible*, i.e. si $\tau : \mathcal{A} \rightarrow \mathcal{B}$ alors pour tout $b \in \mathcal{B}$, le morphisme vérifie $|\tau(b)| \leq 1$.
- Le morphisme $\sigma : \mathcal{A}^* \rightarrow \mathcal{A}^*$ reste prolongeable, mais pour tout $b \in \mathcal{B}$, la longueur $|\sigma(b)|$ n’est plus fixée.

Si le morphisme σ est prolongeable en $a \in \mathcal{A}$, le triplet $T = (\sigma, \tau, a)$ est appelé une *substitution*. Si il existe une substitution $T = (\sigma, \tau, a)$ telle que $\mathbf{x} = \tau(\sigma^{\omega}(a))$, alors il est dit *engendré* par T .

Cela étant, généralisation du Théorème 3.1.2 s’énonce de la façon suivante.

Théorème 3.1.4. *Soit \mathcal{B} un alphabet fini et soit $\mathcal{S} = (L, \mathcal{A}, <_{\mathcal{A}})$ un SNA construit sur un langage régulier. Le mot $\mathbf{x} \in \mathcal{B}^{\mathbb{N}}$ est \mathcal{S} -automatic si et seulement si il est engendré par une substitution T .*

Soit $\mathcal{S} = (L, \mathcal{A}, <_{\mathcal{A}})$ un SNA sur un langage régulier. Nous allons commencer par montrer que si $\mathbf{x} \in \mathcal{B}^{\mathbb{N}}$ est mot \mathcal{S} -automatic, alors il est engendré par une substitution T . Commençons par montrer qu’étant donné un AFD $\mathbb{A} = (Q, q_0, \mathcal{A}, \delta, F)$, nous

pouvons construire un morphisme prolongeable dont le point fixe encode les états atteints dans \mathbb{A} en lisant les mots de \mathcal{A}^* .

Lemme 3.1.5. Soit $\mathcal{A} = \{a_1 < \dots < a_n\}$ un alphabet ordonné et soit $\mathbb{A} = (Q, q_0, \mathcal{A}, \delta, F)$ un AFD complet tel que $\alpha \notin Q$. Le morphisme $\sigma_{\mathbb{A}} : (Q \cup \{\alpha\})^* \rightarrow (Q \cup \{\alpha\})^*$ défini par

$$\begin{cases} \sigma_{\mathbb{A}}(\alpha) &= \alpha q_0, \\ \sigma_{\mathbb{A}}(q) &= \delta(q, a_1) \cdots \delta(q, a_n) \quad \text{pour tout } q \in Q, \end{cases}$$

a pour point fixe le mot $u_{\sigma} \in (Q \cup \{\alpha\})^{\mathbb{N}}$ vérifiant $u_i = \delta(q_0, w_i)$ pour tout $i \in \mathbb{N} \setminus \{0\}$, où w_i est le i -ème mot de \mathcal{A}^* .

Démonstration. Le morphisme $\sigma_{\mathbb{A}}$ est prolongeable en α , et a donc pour point fixe le mot $u_{\sigma} = \sigma_{\mathbb{A}}^{\omega}(\alpha)$. Comme rappelé précédemment, nous savons que

$$\sigma_{\mathbb{A}}^{\omega}(\alpha) = \alpha q_0 \sigma_{\mathbb{A}}(q_0) \sigma_{\mathbb{A}}^2(q_0) \sigma_{\mathbb{A}}^3(q_0) \cdots .$$

Par définition du morphisme $\sigma_{\mathbb{A}}$ cela implique que

$$u_{\sigma} = \alpha q_0 \underbrace{\delta(q_0, a_1) \cdots \delta(q_0, a_n)}_{\sigma_{\mathbb{A}}(q_0)} \underbrace{\delta(q_0, a_1 a_1) \cdots \delta(q_0, a_1 a_n) \cdots \delta(q_0, a_n a_1) \cdots \delta(q_0, a_n a_n)}_{\sigma_{\mathbb{A}}^2(q_0)} \cdots .$$

Étant donné que les premiers mots de \mathcal{A}^* sont

$$\varepsilon \prec a_1 \prec a_2 \prec \cdots \prec a_1 a_1 \prec a_1 a_2 \cdots ,$$

nous pouvons facilement constater grâce au fait que l'alphabet \mathcal{A} est ordonné que $u_i = \delta(q_0, w_i)$ pour tout $i \in \mathbb{N} \setminus \{0\}$ comme souhaité. \square

Remarque 3.1.6. Si l'automate \mathbb{A} défini dans le Lemme 3.1.5 n'est pas complet (i.e. on ne peut pas lire toutes les lettres de \mathcal{A} depuis tous les états de \mathbb{A}), alors pour tout $q \in Q$ nous pouvons définir le morphisme $\sigma_{\mathbb{A}}$ de la même façon en remplaçant $\delta(q, a_i)$ par ε si la transition n'est pas définie. De cette façon, si le langage régulier accepté par \mathbb{A} lorsque tous les états sont finaux est L , alors le mot u_{σ} vérifie $u_i = \delta(q_0, w_i)$ pour tout $i \in \mathbb{N} \setminus \{0\}$ où w_i est le i -ième mot de L .

Grâce au Lemme 3.1.5 et à la proposition suivante, nous allons maintenant établir une substitution engendrant un mot \mathcal{S} -automatique.

Proposition 3.1.7. Soient $\mathbf{u} \in \mathcal{B}^{\mathbb{N}}$ un mot infini et $\mathcal{S} = (L, \mathcal{A}, <_{\mathcal{A}})$ un SNA construit sur un langage L régulier. Si \mathbf{u} est un mot \mathcal{S} -automatique, alors il existe une substitution T engendrant \mathbf{x} .

Démonstration. Soit $\mathbb{A}(L) = (Q, q_0, \delta, \mathcal{A}, F)$ un automate fini ² déterministe acceptant le langage L et soit $\mathbb{A}_S = (Q', q'_0, \delta', \mathcal{A}, \tau, \mathcal{B})$ l'automate fini déterministe avec sortie engendrant le mot \mathbf{x} . Considérons le produit des automates $\mathbb{A}(L)$ et \mathbb{A}_S donné par $\mathbb{P} = (Q \times Q', (q_0, q'_0), \nu, \mathcal{A})$ où $\nu((q, q'), a) = (\delta(q, a), \delta'(q', a))$ pour tout $(q, q') \in Q \times Q'$ et tout $a \in \mathcal{A}$. L'ensemble des états finaux de \mathbb{P} n'est pas donné explicitement car nous n'en aurons pas besoin. Notons $\sigma_{\mathbb{P}} : ((Q \times Q') \cup \{\alpha\})^* \rightarrow ((Q \times Q') \cup \{\alpha\})^*$ le morphisme associé à l'automate \mathbb{P} donné par le Lemme 3.1.5. Considérons en plus le codage effaçant $\mu : ((Q \times Q') \cup \{\alpha\}) \rightarrow \mathcal{B}$ défini de la façon suivante

$$\begin{cases} \mu(\alpha) &= \varepsilon, \\ \mu((q, q')) &= \varepsilon \quad \text{si } q \notin F, \\ &= \tau(q') \quad \text{sinon.} \end{cases}$$

Le morphisme $\sigma_{\mathbb{P}}$ a donc pour point fixe le mot $\sigma_{\mathbb{P}}^{\omega}(\alpha)$ dont les lettres sont la suite d'états obtenue en lisant les mots de \mathcal{A}^* dans l'automate \mathbb{P} . En appliquant à ce mot le codage faible μ , nous supprimons les lettres qui correspondent à la lecture dans \mathbb{P} d'un mot de \mathcal{A}^* n'appartenant pas à L . Les autres lettres sont envoyées sur la sortie correspondante dans l'automate \mathbb{A}_S . La substitution $T = (\sigma_{\mathbb{P}}, \mu, \alpha)$ convient donc. \square

Afin d'associer un AFD à la substitution T , il est parfois intéressant que les morphismes σ et τ vérifient des conditions particulières.

Remarque 3.1.8. [[2]] En réalité, s'il existe une substitution $T = (\sigma, \tau, a)$ engendrant un mot infini \mathbf{x} , alors il existe un codage μ et un morphisme λ non effaçant tels que $\mathbf{x} = \mu(\lambda^{\omega}(a))$. Le codage faible d'une substitution T peut donc toujours être remplacé par un codage μ , tout en gardant un morphisme λ non effaçant.

Montrons maintenant que si \mathbf{x} est engendré par une substitution T alors il existe un SNA \mathcal{S} pour lequel \mathbf{x} est \mathcal{S} -automatique. Commençons par établir comment construire un AFD \mathbb{A}_T associé à une substitution $T = (\sigma, \tau, a)$, où $\sigma : \mathcal{A}^* \rightarrow \mathcal{A}^*$ et $\tau : \mathcal{A} \rightarrow \mathcal{B}$. Par la Remarque 3.1.8, nous pouvons supposer que σ est un morphisme non effaçant et τ un codage. Le AFD $\mathbb{A}_T = (Q, q_0, \mathcal{D}, \delta)$ vérifie

- L'ensemble des états Q sont les lettres de \mathcal{A} ,
- L'état q_0 est a ,
- L'alphabet $\mathcal{D} = \{d_0 < \dots < d_r\}$ où

$$r = \max_{a \in \mathcal{A}} |\sigma(a)| - 1,$$

- Pour tous $b, c \in \mathcal{A}$, s'il existe $0 \leq i < |\sigma(b)|$ tel que $[\sigma(b)]_i = c$ alors $\delta(b, d_i) = c$.
En particulier, il y a une boucle labellisée par la lettre d_0 sur l'état initial a .

2. Cet automate existe grâce à la régularité du langage L .

Pour que l'AFD \mathbb{A}_T soit un AFDS, nous pouvons éventuellement ajouter une fonction de sortie τ' vérifiant $\tau'(a) = \tau(a)$ pour tout $a \in \mathcal{A}$. C'est une des raisons pour lesquelles la Remarque 3.1.8 nous permettant de toujours remplacer un codage faible par un codage τ dans une substitution T est particulièrement utile.

Nous sommes maintenant prêts à montrer que la réciproque de la Proposition 3.1.7 est également vraie.

Proposition 3.1.9. *Si $\mathbf{x} \in \mathcal{B}^{\mathbb{N}}$ est engendré par une substitution $T = (\sigma, \tau, a)$, alors il existe un SNA \mathcal{S} pour lequel \mathbf{x} est \mathcal{S} -automatique.*

Démonstration. Grâce à la Remarque 3.1.8, nous pouvons supposer que $\sigma : \mathcal{A}^* \rightarrow \mathcal{A}^*$ est un morphisme non effaçant et que $\tau : \mathcal{A} \rightarrow \mathcal{B}$ est un codage. Considérons l'automate $\mathbb{A}_T = (Q, q_0, \mathcal{D}, \delta)$ associé à la substitution T et notons $L \subseteq \mathcal{D}^*$ le langage régulier accepté par \mathbb{A}_T lorsque tous ses états sont finaux. Par le Lemme 3.1.5, nous pouvons associer à \mathbb{A}_T un morphisme $\sigma_{\mathbb{A}_T} : (Q \cup \{\alpha\})^* \rightarrow (Q \cup \{\alpha\})^*$ pour $\alpha \notin Q$ (où l'ensemble des états Q sont les lettres de \mathcal{A}). Le point fixe \mathbf{u}_σ de ce morphisme est la suite d'états atteints dans \mathbb{A}_T en lisant les mots de L dans l'ordre radiciel et commençant par α . Cependant, les deux mots infinis $\sigma^\omega(a)$ et $\sigma_{\mathbb{A}_T}^\omega(\alpha)$ sont distincts, l'un commençant par a et l'autre commençant par $\alpha a a$ car la lecture des deux premiers mots ε et d_0 de L mènent dans l'état a . En réalité, les mots $\sigma^\omega(a)$ et $\sigma_{\mathbb{A}_T}^\omega(\alpha)$ coïncident partout sauf en les lettres correspondant à l'état atteint lors de la lecture des mots de L commençant par d_0 dans l'automate \mathbb{A}_T , comme illustré dans l'Exemple 3.1.10. De cette façon, le mot \mathbf{x} est \mathcal{S} -automatique pour l'AFDS défini à partir de \mathbb{A}_T auquel on ajoute une fonction de sortie $\tau' : Q \rightarrow \mathcal{B}$ définie telle que $\tau'(q) = \tau(q)$ pour tout $q \in Q$ et pour le langage régulier $L \setminus d_0 \mathcal{D}^* \subseteq \mathcal{D}^*$. \square

Exemple 3.1.10. Considérons la substitution $T = (\sigma, \tau, a)$ définie telle que

$$\sigma : \begin{cases} a \mapsto ab, \\ b \mapsto bca, \\ c \mapsto cc, \end{cases} \quad \text{et} \quad \tau : \begin{cases} a \mapsto 0, \\ b \mapsto 1, \\ c \mapsto 2. \end{cases}$$

Cette substitution engendre le mot $\mathbf{x} = \tau(\sigma^\omega(a))$. L'automate \mathbb{A}_T associé à la substitution T est représenté à la Figure 3.2. Dans cet exemple, l'alphabet $\mathcal{D} = \{d_0, d_1, d_2\}$ et les états de \mathbb{A}_T sont a, b et c . Le langage $L \subseteq \{d_0, d_1, d_2\}^*$ régulier accepté par l'automate \mathbb{A}_T lorsque tous les états sont accepteurs est composé des premiers mots suivants

$$\varepsilon \prec d_0 \prec d_1 \prec d_0 d_0 \prec d_0 d_1 \prec d_1 d_0 \prec d_1 d_1 \prec d_1 d_2 \prec d_0 d_0 d_0 \prec \dots$$

Par le Lemme 3.1.5, un morphisme $\sigma_{\mathbb{A}_T} : (\{a, b, c\} \cup \alpha)^* \rightarrow (\{a, b, c\} \cup \alpha)^*$ peut être associé à l'AFD \mathbb{A}_T . Ce morphisme a pour point fixe le mot u_σ vérifiant $u_i = \delta(q_0, w_i)$

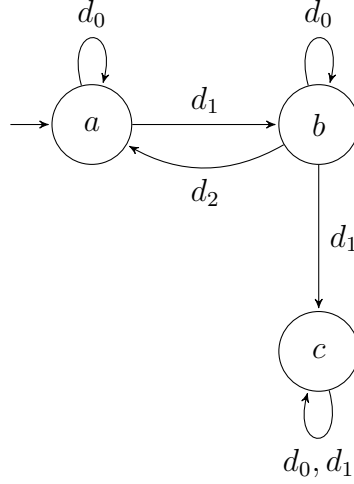


FIGURE 3.2 – L'AFD associé à la substitution T .

pour tout $i \in \mathbb{N} \setminus \{0\}$, où w_i est le i -ème mot de L . Il est défini de la façon suivante,

$$\sigma_{\mathbb{A}_T} : \begin{cases} \alpha \mapsto \alpha a, \\ a \mapsto ab := \sigma(a), \\ b \mapsto bca := \sigma(b), \\ c \mapsto cc := \sigma(c), \end{cases}$$

Les deux mots infinis $\sigma^\omega(a)$ et $\sigma_{\mathbb{A}_T}^\omega(\alpha)$ commencent respectivement par

$$\begin{array}{ccccccccccc} \sigma^\omega(a) & = & a & & b & & & & b & c & a & \cdots \\ \sigma_{\mathbb{A}_T}^\omega(\alpha) & = & \alpha & a & \underbrace{a}_{=\delta(q_0, d_0)} & b & \underbrace{a}_{=\delta(q_0, d_0 d_0)} & \underbrace{b}_{=\delta(q_0, d_0 d_1)} & b & c & a & \cdots, \end{array}$$

et coïncident en les lettres de $\sigma_{\mathbb{A}_T}^\omega(\alpha)$ correspondant à la lecture des mots de L ne commençant pas par d_0 dans l'AFD \mathbb{A}_T . Le mot \mathbf{x} est donc \mathcal{S} -automatique pour le SNA $\mathcal{S} = (L \setminus d_0 \mathcal{D}^*, \mathcal{D}, d_0 < d_1 < d_2)$.

3.1.2 Cas des SNA sur un langage non régulier

Les résultats de la Section 3.1.1 nous ont permis de montrer que les mots \mathcal{S} -automatiques où \mathcal{S} est un SNA sur un langage régulier sont précisément les mots engendrés par une substitution. Or, bien que le cadre du Théorème 3.1.4 soit large, nous n'avons aucune information sur le SNA construit sur un langage non régulier. Dans cette section, nous allons investiguer les SNA construits sur le langage associé à un arbre ayant une signature labellisée purement périodique introduits dans les Sections 1.3.2, 1.3.3 et 2.1.

Comme mentionné à la Remarque 2.2.11, pour tous $p > q > 1$ les langages $L_{\frac{p}{q}}$ sont des cas particuliers de langages associés à un arbre ayant une signature labellisée purement périodique. Ainsi, un mot $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ est dit $\frac{p}{q}$ -automatique s'il est \mathcal{S} -automatique pour un SNA $\mathcal{S} = (L_{\frac{p}{q}}, \mathcal{A}_p, <)$. Les mots $\frac{p}{q}$ -automatiques sont donc un cas particulier de la généralisation du théorème que nous établissons dans cette section.

Exemple 3.1.11. Considérons $(s(n))_{n \geq 0}$ la suite de la somme des chiffres de la représentation en base $\frac{3}{2}$ de n . La représentation des 20 premiers naturels est donnée dans la Table 2.1 et nous avons donc par exemple $s(6) = 2 + 1 + 2 + 0 = 5$. La suite est donnée par

$$(s(n))_{n \geq 0} = 0, 2, 3, 3, 5, 4, 5, 7, 5, 5, 7 \dots$$

Le mot $\mathbf{t}_{\frac{3}{2}}$ qui nous intéresse est $\mathbf{t}_{\frac{3}{2}} = (s(n) \bmod 2)_{n \geq 0}$ dont les premières lettres sont données par

$$\mathbf{t}_{\frac{3}{2}} = 00111011111011 \dots$$

Ce mot est donc défini comme $t_{\frac{3}{2},n} = |\text{rep}_{\frac{3}{2}}(n)|_1 \bmod 2$ pour tout $n \geq 0$ entier et est la généralisation à la base $\frac{3}{2}$ du mot de Thue-Morse \mathbf{t} introduit dans l'Exemple 3.1.1. Le fait de travailler modulo 2 permet que le mot $\mathbf{t}_{\frac{3}{2}}$ ait un alphabet fini. Le mot $\mathbf{t}_{\frac{3}{2}}$ est $\frac{3}{2}$ -automatique et engendré par l'AFDS représenté à la Figure 3.3.

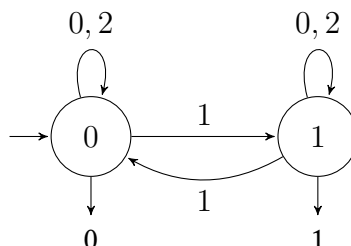


FIGURE 3.3 – L'AFDS engendrant le mot de Thue-Morse généralisé.

Cet AFDS prend en entrée des mots de $L_{\frac{3}{2}}$ et permet d'établir la parité du nombre de 1 dans la représentation en base $\frac{3}{2}$ des naturels. Si la lecture de $\text{rep}_{\frac{3}{2}}(n)$ mène dans l'état 1, cela signifie que le nombre de 1 ayant été lu est impair (et la nombre de 1 est pair si la lecture mène à l'état 0). Les 2 et 0 de la représentation n'apportant aucune information sur le nombre de 1, ils ne permettent pas de se déplacer dans l'automate.

Les mots \mathcal{S} -automatiques pour un SNA construit sur un langage associé à un arbre ayant une signature labellisée purement périodique ne sont plus les mots engendrés par une substitution, mais une généralisation de ceux-ci. Le morphisme σ est remplacé par

un ensemble de r morphismes $(\sigma_0, \dots, \sigma_{r-1})$ et les mots considérés sont l'image par un codage d'un *point fixe alterné* des morphismes $(\sigma_0, \dots, \sigma_{r-1})$. De façon équivalente, ces mots sont l'image par un codage d'un point fixe d'une *substitution par bloc de dimension r* .

Définition 3.1.12. Soit $r \geq 1$ un entier, soit \mathcal{A} un alphabet fini et soient $\sigma_0, \dots, \sigma_{r-1}$ un ensemble de r morphismes uniformes (potentiellement de longueur différente) sur l'alphabet \mathcal{A} . Un mot infini $\mathbf{x} = x_0x_1x_2 \dots \in \mathcal{A}^{\mathbb{N}}$ est un *point fixe alterné*³ de $(\sigma_0, \dots, \sigma_{r-1})$ si

$$\mathbf{x} = \sigma_0(x_0)\sigma_1(x_1) \dots \sigma_{r-1}(x_{r-1})\sigma_0(x_r) \dots \sigma_{i \bmod r}(x_i) \dots .$$

Une *substitution par bloc de dimension r* est une application $g : \mathcal{A}^r \rightarrow \mathcal{A}^*$ qui a un mot $x_0 \dots x_{rn-1} \in \mathcal{A}^*$ de longueur multiple de r associe

$$g(x_0 \dots x_{r-1})g(x_r \dots x_{2r-1}) \dots g(x_{r(n-1)} \dots x_{rn-1}).$$

Si la longueur du mot n'est pas un multiple de r , l'application g laisse inchangée le suffixe du mot. Un mot infini $\mathbf{x} = x_0x_1x_2 \dots \in \mathcal{A}^{\mathbb{N}}$ est un *point fixe de la substitution par bloc de dimension r* $g : \mathcal{A}^r \rightarrow \mathcal{A}^*$ si

$$\mathbf{x} = g(x_0 \dots x_{r-1})g(x_r \dots x_{2r-1}) \dots .$$

Remarque 3.1.13. Les deux notions introduites dans la Définition 3.1.12 sont équivalentes et nous pouvons maintenant montrer que un mot $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ est un point fixe alterné des r morphismes uniformes $(\sigma_0, \dots, \sigma_{r-1})$ sur \mathcal{A}^* si et seulement si c'est un point fixe d'une substitution par bloc de dimension r $g : \mathcal{A}^r \rightarrow \mathcal{A}^*$.

Démonstration. Pour tout mot $a_0 \dots a_{r-1} \in \mathcal{A}^r$, on peut définir la substitution par bloc de dimension r donnée par

$$g : \mathcal{A}^r \rightarrow \mathcal{A}^*, \\ a_0 \dots a_{r-1} \mapsto g(a_0 \dots a_{r-1}) = \sigma_0(a_0) \dots \sigma_{r-1}(a_{r-1}).$$

□

Exemple 3.1.14. Considérons les morphismes (σ_0, σ_1) définis comme

$$\sigma_0 = \begin{cases} 0 & \mapsto 00 \\ 1 & \mapsto 11 \end{cases} \quad \text{et} \quad \sigma_1 = \begin{cases} 0 & \mapsto 1 \\ 1 & \mapsto 0. \end{cases}$$

3. Cette notion de point fixe généralise la notion de point fixe d'un morphisme σ prolongeable. En effet, si \mathbf{x} est un point fixe du morphisme σ , alors

$$\mathbf{x} = \sigma(\mathbf{x}) = \sigma(x_0)\sigma(x_1)\sigma(x_2) \dots .$$

Le point fixe alterné de (σ_0, σ_1) n'est autre que $\mathbf{t}_{\frac{3}{2}}$. Par la Remarque 3.1.13, le mot $\mathbf{t}_{\frac{3}{2}}$ est également un point fixe de la substitution par bloc de dimension 2 telle que $g : \{0, 1\}^2 \rightarrow \{0, 1\}^*$ définie par

$$g : \begin{cases} 00 \mapsto \sigma_0(0)\sigma_1(0) = 001 \\ 01 \mapsto \sigma_0(0)\sigma_1(1) = 000 \\ 10 \mapsto \sigma_0(1)\sigma_1(0) = 111 \\ 11 \mapsto \sigma_0(1)\sigma_1(1) = 110. \end{cases} \quad (3.2)$$

Les Propositions 3.1.18 et 3.1.20 nous permettent de démontrer formellement que $\mathbf{t}_{\frac{3}{2}}$ est un point fixe alterné des morphismes (σ_0, σ_1) . Nous pouvons également le montrer intuitivement sans leur aide en procédant de façon similaire à [12]. Le mot $(s(n))_{n \geq 0}$ est défini de façon à ce que pour tout $n \geq 0$, $s(n)$ est la somme des chiffres labellisant le chemin entre la racine et le n -ième noeud de l'arbre représenté à la Figure 1.6. Au vu de la signature de l'arbre, les noeuds ont successivement un ou deux fils. Si un noeud m à deux fils m_1 et m_2 , alors $s(m_1) = s(m) + 0$ et $s(m_2) = s(m) + 2$ (similairement pour un seul fils). Si un mot infini \mathbf{x} est un point fixe d'une substitution par blocs de dimension 2 notée h dont toutes les images sont de longueur constante, alors par définition

$$\mathbf{x} = h(x_0x_1)h(x_2x_3) \cdots$$

Le mot \mathbf{x} est donc la concaténation des images de blocs de dimension 2 par h . Les blocs de dimension 2 pris en argument sont des blocs de la forme $x_{2i}x_{2i+1}$ dont l'image est de la forme $x_{3i}x_{3i+1}x_{3i+2}$. Dans l'arbre de la Figure 1.6, les noeuds de la forme $2i$ pour i entier, ont toujours deux fils. On peut donc construire une substitution par bloc de dimension 2 notée f qui a pour point fixe de mot $(s(n))_{n \geq 0}$ en simulant le passage d'un bloc de deux noeud à leurs fils de la façon suivante

$$a, b \rightarrow a + 0, a + 2, b + 1 \quad \text{pour } a, b \in \mathbb{N}.$$

Le mot $\mathbf{t}_{\frac{3}{2}}$ étant obtenu à partir de $(s(n))_{n \geq 0}$ en prenant chaque lettre modulo 2, il est le point fixe de la substitution par bloc de dimension 2 notée g et définie comme dans (3.2) obtenue en considérant f modulo 2.

Cela étant, nous pouvons maintenant énoncer le théorème qui généralise le théorème de Cobham 3.1.2 aux SNA construits sur un langage associé à un arbre ayant une signature labellisée purement périodique.

Théorème 3.1.15. *Soient \mathcal{A}, \mathcal{B} des alphabets finis. Un mot infini sur l'alphabet \mathcal{B} est l'image par un codage $\tau : \mathcal{A} \rightarrow \mathcal{B}$ d'un point fixe alterné de morphismes uniformes sur \mathcal{A} (qui ne sont pas nécessairement de même longueur) si et seulement si il est \mathcal{S} -automatique pour un SNA \mathcal{S} construit sur le langage associé à un arbre ayant une signature labellisée purement périodique.*

La démonstration du Théorème 3.1.15 est décomposée en deux propositions traitant respectivement la condition nécessaire et la condition suffisante. Avant ces deux propositions, nous avons encore besoin d'un résultat nous permettant de connaître la forme du point fixe alterné d'un ensemble de r morphismes.

Lemme 3.1.16. *Soit $r \geq 1$ un entier et soit \mathcal{A} un alphabet fini et soient $\sigma_0, \dots, \sigma_{r-1}$ r morphismes sur \mathcal{A}^* . Soit $\mathbf{x} = x_0x_1x_2 \dots$ un point fixe alterné de $(\sigma_0, \dots, \sigma_{r-1})$. Pour tout $m \geq 0$, nous avons*

$$\sigma_{m \bmod r}(x_m) = x_i \cdots x_{i+|\sigma_{m \bmod r}(x_m)|-1}$$

$$\text{où } i = \sum_{j=0}^{m-1} |\sigma_{j \bmod r}(x_j)|.$$

Démonstration. Soit $m \geq 0$. En utilisant la définition d'un point fixe alterné, on peut réécrire

$$\mathbf{x} = u\sigma_{m \bmod r}(x_m)\sigma_{m+1 \bmod r}(x_{m+1}) \cdots$$

où u est défini comme

$$u = \sigma_0(x_0)\sigma_1(x_1) \cdots \sigma_{r-1}(x_{r-1})\sigma_0(x_r) \cdots \sigma_{(m-1) \bmod r}(x_{m-1}).$$

Le préfixe u de w étant de longueur $\sum_{j=0}^{m-1} |\sigma_{j \bmod r}(x_j)|$, cela conclut la preuve. \square

Remarque 3.1.17. Si on considère \mathbf{x} le point fixe du morphisme k -uniforme $\sigma : \mathcal{A}^* \rightarrow \mathcal{A}^*$, le Lemme 3.1.16 se simplifie grandement. En effet au vu du fait que $|\sigma(a)| = k$ pour tout $a \in \mathcal{A}$ et que $\sigma(\mathbf{x}) = \mathbf{x}$, l'image de x_m est $\sigma(x_m) = x_{mk} \cdots x_{(m+1)k-1}$ car $i = \sum_{j=0}^{m-1} |\sigma(x_j)| = km$ dans ces conditions.

Commençons par la condition nécessaire du Théorème 3.1.15.

Proposition 3.1.18. *Soit $r \geq 1$ un entier et soit \mathcal{A} un alphabet fini. Soient $\sigma_0, \dots, \sigma_{r-1} : \mathcal{A}^* \rightarrow \mathcal{A}^*$ r morphismes de longueur respective l_0, \dots, l_{r-1} , tels que σ_0 est prolongeable en une lettre $a \in \mathcal{A}$. Soit $\mathbf{x} = x_0x_1x_2 \cdots \in \mathcal{A}^{\mathbb{N}}$ un point fixe alterné de $(\sigma_0, \dots, \sigma_{r-1})$ commençant par a . Considérons le langage $L(\mathbf{s})$ associé au i -arbre ayant pour signature labellisée purement périodique*

$$\mathbf{s} = \left(0 \cdots (l_0 - 1), l_0(l_0 + 1) \cdots (l_0 + l_1 - 1), \dots, \left(\sum_{j < r-1} l_j \right) \cdots \left(\left(\sum_{j < r} l_j \right) - 1 \right) \right)^\omega,$$

formée d'entiers positifs consécutifs. Soit \mathbb{A} l'AFD où,

- \mathcal{A} est l'ensemble des états,
- a est l'état initial
- $\mathcal{B} = \{0, \dots, \sum_{j < r} l_j - 1\}$ est l'alphabet de l'automate,

- la fonction de transition est définie comme suit : Pour tout $i \in \mathcal{B}$, il existe un unique $j_i \geq 0$ et un unique $t_i \geq 0$ tels que $i = \sum_{k \leq j_i - 1} l_k + t_i$ avec $t_i < l_{j_i}$. On pose

$$\delta(b, i) = [\sigma_{j_i}(b)]_{t_i}, \quad \forall b \in \mathcal{A}.$$

Le mot \mathbf{x} est alors la suite d'états atteints dans \mathbb{A} en lisant les mots de $L(\mathbf{s})$ dans l'ordre radiciel, i.e., pour tout $n \geq 0$, $x_n = \delta(a, \text{rep}_{\mathcal{S}}(n))$ avec $\mathcal{S} = (L(\mathbf{s}), \mathcal{B}, <)$.

Dans le cas où le mot \mathbf{x} est le point fixe d'un unique morphisme σ_0 prolongeable en $a \in \mathcal{A}$ et de longueur k , la Proposition 3.1.18 nous permet de retomber sur une partie du théorème de Cobham concernant les suites k -automatiques. En effet, si σ_0 est l'unique morphisme, la signature \mathbf{s} se réduit à

$$\mathbf{s} = (01 \cdots (k-1))^\omega,$$

et le langage associé $L(\mathbf{s})$ n'est autre que \mathcal{A}_k^* . L'AFD de l'énoncé a dans ce cas $\mathcal{B} = \{0, \dots, k-1\} := \mathcal{A}_k$ comme alphabet. Pour tout $i \in \mathcal{B} := \mathcal{A}_k$, nous avons $j_i = 0$ car la signature \mathbf{s} n'est composée que d'un seul mot. La somme $\sum_{k \leq j_i - 1} l_k$ définissant les transitions dans l'énoncé permet de se placer dans le mot de la signature contenant la lettre i (le mot contenant i dans la signature est unique au vu de sa définition), et la signature ne se composant que d'un seul mot dans notre cas, la somme n'a plus d'intérêt. Nous avons donc $i = t_i$ et la fonction de transition est définie telle que

$$\delta(b, i) = [\sigma_0(b)]_i \quad \text{pour tout } b \in \mathcal{A}.$$

L'automate décrit dans le cas où le mot \mathbf{x} est le point fixe d'un unique morphisme σ_0 est précisément celui construit dans le cadre de la démonstration du théorème de Cobham pour les suites k -automatiques. (Voir [Théorème 6.3.2,[2]])

Intuitivement, l'AFD \mathbb{A} est défini de façon à se déplacer correctement dans l'arbre $T(\mathbf{s})$, tout en ayant les bonnes décorations sur les noeuds de l'arbre. Dans l'arbre $T(\mathbf{s})$ représenté à la Figure 3.4, nous avons $l_0 - 1 \equiv n \pmod{r}$.

La signature \mathbf{s} de l'arbre $T(\mathbf{s})$ est construite de façon particulière :

- La longueur des mots de \mathbf{s} est exactement celle des morphismes $\sigma_0, \dots, \sigma_{r-1}$. L'arbre $T(\mathbf{s})$ étant décoré par le mot \mathbf{x} qui est le point fixe alterné des morphismes $(\sigma_0, \dots, \sigma_{r-1})$, nous savons par le Lemme 3.1.16 que les noeuds allant de i à $i + |\sigma_{m \bmod r}(x_m)| - 1$ pour $i = \sum_{n=0}^{m-1} \sigma_{n \bmod r}(x_n)$ doivent être décorés par les lettres composant $\sigma_{m \bmod r}(x_m)$. La longueur des mots de la signature permet alors que les noeuds labellisés par $\sigma_{m \bmod r}(x_m)$ soient les fils du noeud labellisé par x_m . En particulier, étant donné que x_0 décore le noeud 0 dans l'arbre, nous savons que les noeuds allant de 0 à $l_0 - 1$ doivent être décorés par $\sigma_0(x_0)$. Le fait que la racine ait l_0 fils permet alors que les noeuds décorés par $\sigma_1(x_1)$ soient les fils du noeud décoré par x_1 .

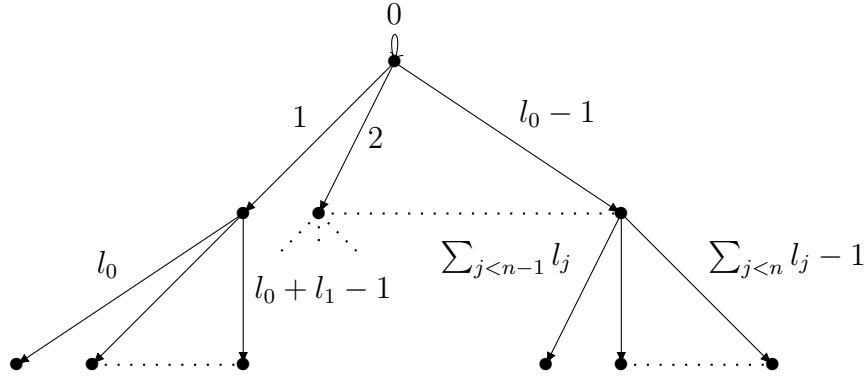


FIGURE 3.4 - Arbre associé à la signature labellisée purement périodique $\mathbf{s} = \left(0 \cdots (l_0 - 1), l_0 \cdots l_0 + l_1 - 1, \dots, \sum_{j < r-1} l_j \cdots \left(\left(\sum_{j < r} l_j\right) - 1\right)\right)^\omega$.

- Chacun des mots de \mathbf{s} appartient à $\text{inc}(\mathcal{B}^*)^4$ et les mots de la signature sont eux mêmes ordonnés entre eux. Ils sont donc tous composés de lettres différentes et toute lettre de \mathcal{B} apparait dans un unique mot de la signature labellisée purement périodique. Cela permet de définir univoquement la fonction de transition de \mathbb{A} .

Démonstration. Procédons par récurrence sur $n \geq 0$. Dans le cas où $n = 0$, nous avons par hypothèse $x_0 = a = \delta(a, \varepsilon)$ car $\text{rep}_{\mathcal{S}}(0) = \varepsilon$. Soit $n \geq 1$. Supposons maintenant que la propriété est vérifiée pour toutes les valeurs inférieures à n et montrons-le pour n .

Notons $\text{rep}_{\mathcal{S}}(n) = a_l \cdots a_0$. Nous avons donc dans le i -arbre associé à la signature périodique \mathbf{s} un chemin démarrant de la racine et allant jusqu'au n -ième noeud qui est labellisé par le mot $a_l \cdots a_0$. Le langage $L(\mathbf{s})$ associé au i -arbre généré par \mathbf{s} étant clos par préfixe, il existe un entier m tel que $m < n$ et tel que $\text{rep}_{\mathcal{S}}(m) = a_l \cdots a_1$. Posons $j = m \bmod r$. Dans le i -arbre généré par la signature \mathbf{s} , lire le mot $a_l \cdots a_1$ depuis la racine conduit à un noeud ayant l_j fils. Comme illustrer dans la Figure 3.5, ces fils sont respectivement atteints par des arêtes labellisées par

$$\sum_{k \leq j-1} l_k, \left(\sum_{k \leq j-1} l_k\right) + 1, \dots, \left(\sum_{k \leq j} l_k\right) - 1.$$

Les l_j étant des entiers positifs pour tout $j \in \{0, \dots, r-1\}$, le mot

$$\left(\sum_{k \leq j-1} l_k\right) \left(\left(\sum_{k \leq j-1} l_k\right) + 1\right) \cdots \left(\left(\sum_{k \leq j} l_k\right) - 1\right)$$

4. Cette notation est définie dans la Section 1.1.

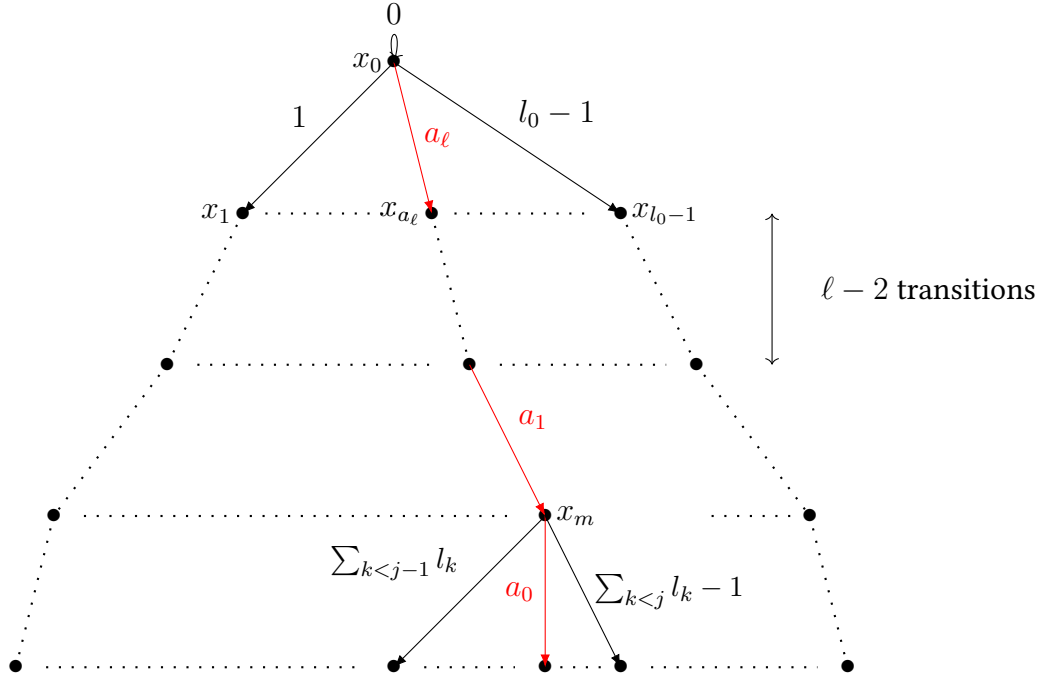


FIGURE 3.5 – Illustration des fils du noeud x_m .

appartient à $\text{inc}(\mathcal{B}^*)$, et la lettre a_0 n'y apparait qu'une unique fois. De plus, les mots de \mathcal{s} étant eux-même ordonnés, la lettre a_0 n'apparait dans aucun autre mot de \mathcal{s} . Supposons que $a_0 = \left(\sum_{k \leq j-1} l_k\right) + t$ pour $t \in \{0, \dots, l_j - 1\}$. En remarquant que les mots de \mathcal{s} sont de longueur respective l_0, \dots, l_{r-1} , nous avons par construction du i -arbre

$$n = \sum_{\substack{v \in L(\mathcal{s}) \\ v < \text{rep}_{\mathcal{S}}(m)}} \text{deg}(v) + t = \sum_{t=0}^{m-1} l_{i \bmod r} + t. \quad (3.3)$$

Par hypothèse de récurrence, nous obtenons

$$\delta(a, \text{rep}_{\mathcal{S}}(n)) = \delta(\delta(a, \text{rep}_{\mathcal{S}}(m)), a_0) = \delta(x_m, a_0)$$

et par définition de la fonction de transition, $\delta(x_m, a_0) = [\sigma_j(x_m)]_t = [\sigma_{m \bmod r}(x_m)]_t$. Nous pouvons donc conclure que c'est exactement x_m en utilisant le Lemme 3.1.16 et l'Égalité (3.3). \square

Soit $\mathbf{x} \in \mathcal{A}^{\mathbb{N}}$ vérifiant les conditions de la Proposition 3.1.18 et soit $\mathbf{y} = \tau(\mathbf{x})$ où $\tau : \mathcal{A} \rightarrow \mathcal{B}$ est un codage. Dans ce cas, \mathbf{y} est l'image par un codage d'un point fixe alterné des morphismes $(\sigma_0, \dots, \sigma_{r-1})$. Pour montrer que $\mathbf{y} \in \mathcal{B}^{\mathbb{N}}$ est \mathcal{S} -automatique

pour un SNA \mathcal{S} défini sur un langage associé à un arbre ayant une signature labellisée purement périodique, il suffit donc d'établir une fonction de sortie τ' sur l'AFD \mathbb{A} défini dans la Proposition 3.1.18 vérifiant $\tau'(a) = \tau(a)$ pour tout $a \in \mathcal{A}$. Ceci conclut la preuve de la condition nécessaire du Théorème 3.1.15.

Exemple 3.1.19. Considérons le mot de Thue-Morse généralisé $\mathbf{t}_{\frac{3}{2}}$ qui est le point fixe alterné des morphismes (σ_0, σ_1) définis dans l'Exemple 3.1.14. La Proposition 3.1.18 nous apprend que $\mathbf{t}_{\frac{3}{2}}$ est \mathcal{S} -automatique pour $\mathcal{S} = (L(\mathbf{s}), \mathcal{A}_2, <)$ où $\mathbf{s} = (01, 2)^\omega$. Le langage $L(\mathbf{s})$ est l'image du langage $L_{\frac{3}{2}}$ par le codage h défini par

$$h : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 2 \\ 2 \mapsto 1. \end{cases}$$

et l'AFDS pour lequel $\mathbf{t}_{\frac{3}{2}}$ est \mathcal{S} -automatique est celui représenté à la Figure 3.3 où les transitions sont remplacées par leur image par le codage h .

Proposition 3.1.20. Soit $r \geq 1$ un entier et soit \mathcal{A} un alphabet fini de chiffres. Soient w_0, \dots, w_{r-1} des mots non vides de $\text{inc}(\mathcal{A}^*)$. Considérons le langage $L(\mathbf{s})$ du i -arbre généré par la signature purement périodique $\mathbf{s} = (w_0, w_1, \dots, w_{r-1})^\omega$. Soit $\mathbb{A} = (Q, q_0, \mathcal{A}, \delta)$ un AFD. Pour $i \in \{0, \dots, r-1\}$, on définit r morphismes de la façon suivante

$$\sigma_i : Q \rightarrow Q^{|w_i|}, q \mapsto \delta(q, w_{i,0})\delta(q, w_{i,1}) \cdots \delta(q, w_{i,|w_i|-1}),$$

où $w_{i,j}$ est la j -ième lettre de w_i . Le point fixe alterné $\mathbf{x} = x_0x_1x_2 \cdots$ de $(\sigma_0, \sigma_1, \dots, \sigma_{r-1})$ commençant par q_0 est la suite d'états atteints dans \mathbb{A} quand on lit les mots de $L(\mathbf{s})$ dans l'ordre radiciel, i.e., pour tout $n \geq 0$, $x_n = \delta(q_0, \text{rep}_{\mathcal{S}}(n))$ avec $\mathcal{S} = (L(\mathbf{s}), \mathcal{A}, <)$.

Comme dans la Section 3.1.1 concernant le cas des SNA construits sur un langage régulier, nous allons procéder en construisant un morphisme particulier (qui est ici un ensemble de morphismes) dont le point fixe alterné \mathbf{x} a pour n -ième lettre l'état atteint lors de la lecture du n -ième mot du langage associé à $T(\mathbf{s})$. Avant de passer à la démonstration, commençons par quelques observations.

Si la signature labellisée purement périodique est de la forme $\mathbf{s} = (w_0)^\omega$, la Proposition 3.1.20 correspond à une partie de la condition nécessaire du théorème de Cobham 3.1.2 pour les suites k -automatiques. Plus précisément, si $w_0 = 012 \cdots (k-1)$, l'arbre $T(\mathbf{s})$ associé est représenté à la Figure 3.6. Cet arbre n'est autre que $T(L_k)$ auquel une boucle labellisée par 0 a été ajoutée sur la racine.

L'ensemble des r morphismes $(\sigma_0, \dots, \sigma_{r-1})$ est résumé à un unique morphisme σ_0 que nous notons σ . Le point fixe du morphisme σ commençant par q_0 sera $\mathbf{x} = \sigma^\omega(q_0)$

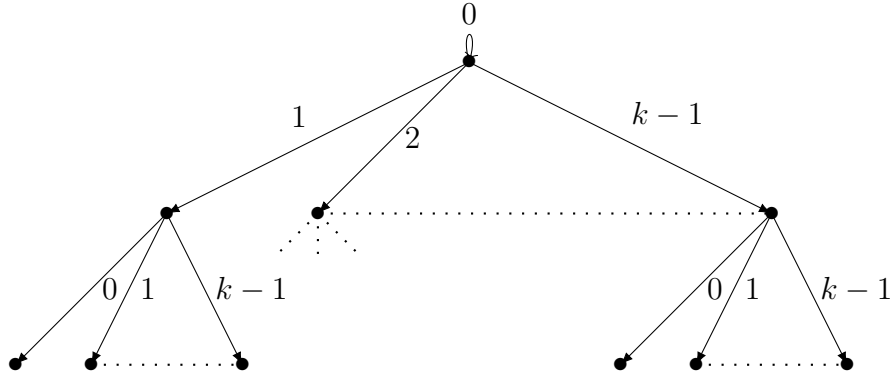


FIGURE 3.6 – Arbre associé à la signature labellisée purement périodique $(01 \cdots (k-1))^\omega$.

et par l'Egalité (3.1)

$$\begin{aligned} \mathbf{x} &= q_0 \delta(q_0, 1) \cdots \delta(q_0, k-1) \sigma \left(\delta(q_0, 1) \cdots \delta(q_0, k-1) \right) \sigma^2 \left(\delta(q_0, 1) \cdots \delta(q_0, k-1) \right) \cdots \\ &= q_0 \delta(q_0, 1) \cdots \delta(q_0, k-1) \delta(q_0, 10) \delta(q_0, 1(k-1)) \cdots \delta(q_0, (k-1)0) \cdots \delta(q_0, (k-1)(k-1)) \cdots \end{aligned}$$

Le mot \mathbf{x} est bien la suite d'états atteints dans \mathbb{A} quand on lit les mots de $L(\mathbf{s})$ car dans ce cas $L(\mathbf{s}) = \mathcal{A}_k^*$. Au vu de la forme de l'arbre, la lecture de tout mot w non vide de $L(\mathbf{s})$ mène à un noeud ayant k fils, auxquels il est relié par des arêtes labellisées par les lettres $0, \dots, k-1$. Toutes les lettres des \mathcal{A}_k peuvent donc être lues depuis le noeud atteint en ayant lu w .

Dans le cas d'une signature labellisée purement périodique composée de plus d'un mot, la situation se complique.

Par définition, l'arbre associé à la signature labellisée purement périodique $\mathbf{s} = (w_0, \dots, w_{r-1})^\omega$ est $T(\mathbf{s})$ qui est représenté à la Figure 3.7.

Pour rappel, la lecture du n -ième mot du langage associé à l'arbre $T(\mathbf{s})$ correspond à la lecture des labels sur les arêtes empruntées entre la racine et le n -ième noeud de l'arbre. La différence avec le cas où la signature labellisée purement périodique est composée d'un seul mot est que la lecture d'un mot peut mener à r noeuds différents, qui ont un nombre de fils lié à la valeur du noeud modulo r . Plus précisément, si le noeud atteint est le n -ième, son nombre de fils sera $|w_{n \bmod r}|$ et les arêtes entre le noeud et ses fils seront labellisées par les lettres de $w_{n \bmod r}$. De ce fait, les lettres pouvant être lues dans l'arbre $T(\mathbf{s})$ dépendent du noeud atteint lors de la lecture du mot w . Cela explique la raison pour laquelle r morphismes distincts doivent être définis à la place d'un seul. Pour tout $i \in \{0, \dots, r-1\}$, le morphisme σ_i correspond au facteur de hauteur de 1 de $T(\mathbf{s})$ représenté à la Figure 3.8.

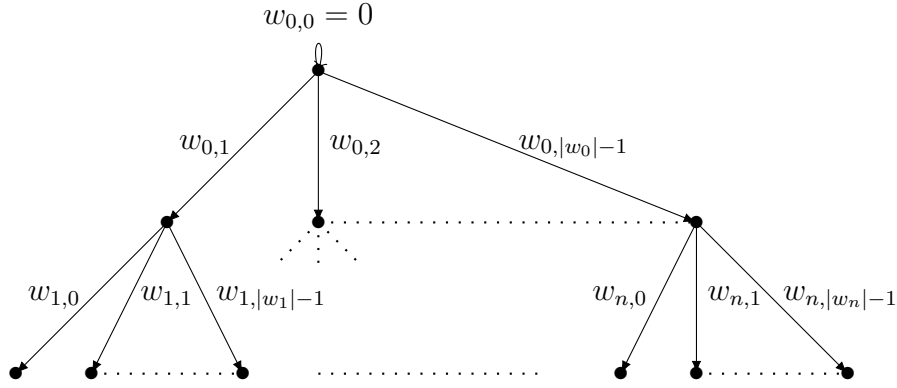


FIGURE 3.7 – Premiers niveaux de l'arbre $T(\mathbf{s})$ associé à la signature labellisée purement périodique $\mathbf{s} = (w_0, \dots, w_{r-1})$.

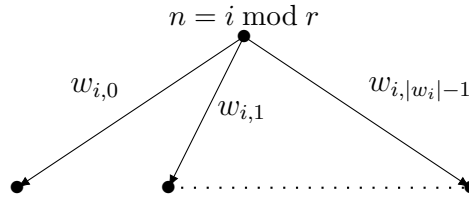


FIGURE 3.8 – Arbre associé au morphisme σ_i .

Par définition, le point fixe alterné \mathbf{x} des r morphismes $(\sigma_0, \dots, \sigma_{r-1})$ commençant par q_0 est donc de la forme

$$\mathbf{x} = \underbrace{q_0 \delta(q_0, w_{0,1}) \delta(q_0, w_{0,|w_0|-1})}_{\sigma_0(q_0)} \underbrace{\delta(q_0, w_{0,1}w_{1,0}) \cdots \delta(q_0, w_{0,1}w_{1,|w_1|-1})}_{\sigma_1([\sigma_0(q_0)]_1)} \underbrace{\delta(q_0, w_{0,2}w_{2,0}) \cdots \delta(q_0, w_{0,2}w_{2,|w_2|-1})}_{\sigma_2([\sigma_0(q_0)]_2)} \cdots$$

Le morphisme σ_0 permet de passer de l'étage 0 à l'étage 1 de l'arbre en empruntant les arêtes labellisées par le mot w_0 . La lecture de $w_{0,1}$ nous mène dans un sous-arbre dont la racine est un noeud qui vaut 1 modulo r , ce qui implique qu'il faut appliquer le morphisme σ_1 pour continuer de lire les mots de l'arbre en ayant déjà lu $w_{0,1}$.

Passons maintenant à la démonstration du résultat.

Démonstration. Au vu de la définition d'une signature labellisée, on peut supposer que w_0 s'écrit $w_0 = 0u$ avec $u \in \mathcal{A}^+$. Si ce n'est pas le cas, il suffit de renommer les lettres de l'alphabet \mathcal{A} .

Procédons par récurrence sur $n \geq 0$. Le cas $n = 0$ est direct, on a que $x_0 = \delta(q_0, \text{rep}_S(0)) = \delta(q_0, \varepsilon) = q_0$, ce qui est suffisant au vu de l'énoncé.

Supposons que la propriété soit vérifiée pour toutes les valeurs inférieures à n et montrons qu'elle reste valide pour n . Notons $\text{rep}_S(n) = a_l \cdots a_1 a_0$. Par définition de l'arbre associé à $L(s)$, cela signifie que les arêtes parcourues de la racine au n -ième noeud de l' i -arbre forment un chemin labellisé par le mot $a_l \cdots a_0$. Comme $L(s)$ est un langage clos par préfixe, le mot $a_l \cdots a_1$ appartient à $L(s)$ et il existe un entier m tel que $m < n$ et $\text{rep}_S(m) = a_l \cdots a_1$. Posons $i = m \bmod r$. Par construction du i -arbre associé signature périodique s , le parcours du chemin labellisé par $a_l \cdots a_1$ nous permet d'atteindre le m -ième noeud de l'arbre, qui possède $|w_i|$ fils. Sachant qu'une des $|w_i|$ arêtes doit être empruntée pour passer du m -ième noeud au n -ième noeud de l'arbre, une des lettres de w_i doit être a_0 . De plus étant donné que $w_i \in \text{inc}(\mathcal{A}^*)$, a_0 n'apparaît qu'une fois dans w_i . Supposons que $a_0 = w_{i,j}$ pour un $j \in \{0, \dots, |w_i| - 1\}$. On a donc,

$$n = \sum_{k=0}^{m-1} |w_{k \bmod r}| + j. \quad (3.4)$$

Par hypothèse de récurrence, on a

$$\delta(q_0, \text{rep}_S(n)) = \delta(\delta(q_0, \text{rep}_S(m)), a_0) = \delta(x_m, a_0)$$

et par définition de σ_i nous obtenons,

$$\delta(x_m, a_0) = [\sigma_i(x_m)]_j = [\sigma_{m \bmod r}(x_m)]_j.$$

Nous pouvons dès lors conclure que c'est exactement x_n en se servant du Lemme 3.1.16 et de l'Égalité (3.4). \square

Exemple 3.1.21. Considérons le mot de Thue-Morse généralisé $\mathbf{t}_{\frac{3}{2}}$. Dans la Proposition 3.4, le fait de considérer l'AFDS défini à la Figure 3.3 et la signature labellisée $s = (02, 1)^\omega$ ou de considérer l'AFDS de la Figure 3.3 où les labels des transitions sont remplacés par leur image par le codage h de l'Exemple 3.1.19 et la signature $(01, 2)^\omega$ donne les mêmes morphismes (σ_0, σ_1) ,

$$\sigma_0 = \begin{cases} 0 & \mapsto 00 \\ 1 & \mapsto 11 \end{cases} \quad \text{et} \quad \sigma_1 = \begin{cases} 0 & \mapsto 1 \\ 1 & \mapsto 0. \end{cases}$$

L'important ici ne sont donc pas les chiffres composant la signature labellisée purement périodique, mais bien la forme de l'arbre associé à cette signature.

La généralisation du théorème de Cobham aux suites $\frac{p}{q}$ -automatiques découle des deux propositions précédentes, pour un SNA $\mathcal{S} = (L_{\frac{p}{q}}, \mathcal{A}_p, <)$. Nous avons un résultat supplémentaire dans ce cas d'intérêt. L'arbre $T(L_{\frac{p}{q}})$ à une signature labellisée purement périodique de la forme $(w_0, \dots, w_{q-1})^\omega$ avec $\sum_{i=0}^{q-1} |w_i| = p$ et $w_i \in \mathcal{A}_p^*$ pour tout i .

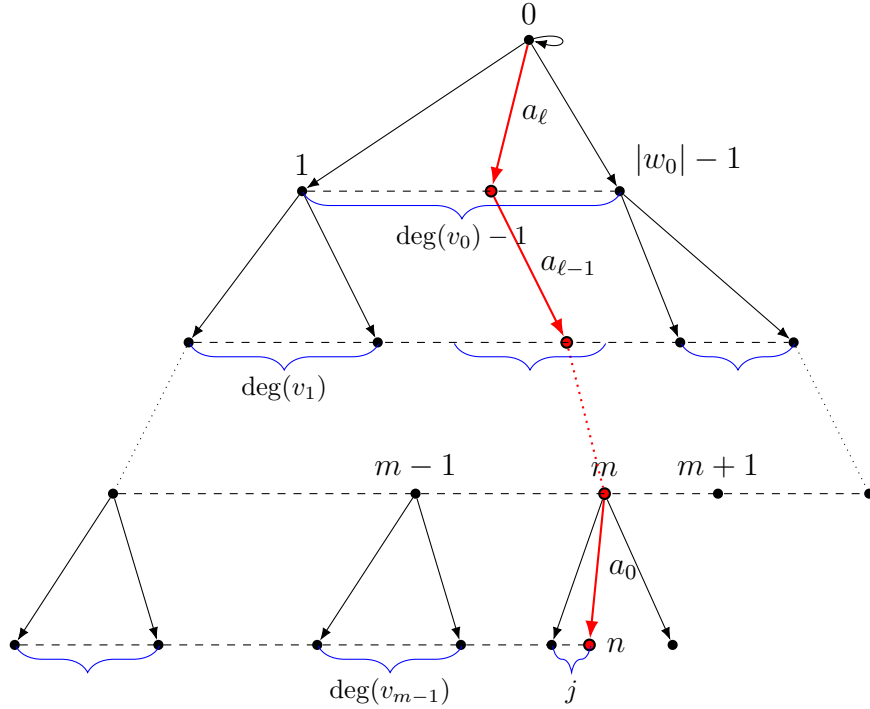


FIGURE 3.9 – Illustration de l'Equation (3.4).

Corollaire 3.1.22. *Dans le cas où la suite est $\frac{p}{q}$ -automatique, la suite est l'image par un codage d'un point fixe d'une substitution par bloc de dimension q dont les images sont toutes de longueur p .*

Démonstration. Notons $(w_0, \dots, w_{q-1})^\omega$ la signature purement périodique labellisée associée à l'arbre $T(L_{\frac{p}{q}})$. La Proposition 3.1.20 nous donne q morphismes $|w_i|$ -uniformes pour $i \in \{0, \dots, q-1\}$. Par la Remarque 3.1.13, le point fixe alterné de $(\sigma_0, \dots, \sigma_{q-1})$ est un point fixe d'une substitution par bloc de dimension q notée g , telle que pour tout mot de longueur q dénoté par $a_0 \dots a_{q-1}$,

$$|g(a_0 \dots a_{q-1})| = |\sigma_0(a_0)\sigma_1(a_1) \dots \sigma_{q-1}(a_{q-1})| = \sum_{i=0}^{q-1} |w_i| = p.$$

□

3.2 Caractérisation des suites \mathcal{S} -automatiques

Dans cette section, nous allons établir une caractérisation de la \mathcal{S} -automaticité d'une suite en exploitant le lien entre les SNA $\mathcal{S} = (L, \mathcal{A}, <_{\mathcal{A}})$ associés à un langage clos par préfixe et l'arbre $T(L)$ décoré par le mot $\mathbf{x} \in \mathcal{B}^{\mathbb{N}}$ (noté $T_{\mathbf{x}}(L)$). Plus

précisément, la caractérisation que nous allons établir lie l'automaticité et le nombre de sous-arbres de $T(L)$.

Considérons à partir de maintenant un langage L clos par préfixe sur un alphabet $(\mathcal{A}, <_{\mathcal{A}})$ et un mot $\mathbf{x} = x_0x_1x_2\cdots \in \mathcal{B}^{\mathbb{N}}$ où \mathcal{B} est un alphabet fini décorant $T(L)$, qui est noté T dans cette section par souci de clarté. Pour rappel, cet arbre a son n -ième noeud décoré par la lettre $x_n \in \mathcal{B}$ et le chemin parcouru de la racine au n -ième noeud est labellisé par le n -ième mot de L . Soit $\mathcal{S} = (L, \mathcal{A}, <_{\mathcal{A}})$ un SNA. Si $\text{rep}_{\mathcal{S}}(n) = w \in L$, alors $x_n = x_{\text{val}_{\mathcal{S}}(w)}$. En particulier, si $w, w' \in L$ et $h \geq 0$ entier, deux sous-arbres $T[w, h], T[w', h]$ sont isomorphes si $x_{\text{val}_{\mathcal{S}}(wu)} = x_{\text{val}_{\mathcal{S}}(w'u)}$ pour tout $u \in \text{dom}(T[w, h]) = \text{dom}(T[w', h])$. On note également

$$F_h = \{T[w, h] \mid w \in L\}$$

l'ensemble des facteurs de hauteur h apparaissant dans T . L'arbre T est *rationnel* si il a un nombre fini de suffixes. En particulier l'ensemble $F_h^{\infty} \subset F_h$ est l'ensemble de facteurs de hauteur h de T y apparaissant infiniment souvent. De plus, pour tout a convenable labellisant les arêtes T , nous notons $F_{h,a}^{\infty} \subset F_h^{\infty}$ l'ensemble des sous-arbres de T de hauteur h , apparaissant infiniment souvent dans T et ayant a comme lettre labellisant la première arête entre le niveau $h - 1$ et h . Cela est identique à dire que le premier mot du domaine du sous-arbre considéré finit par a .

Nous allons établir un résultat caractérisant la \mathcal{S} -automaticité dans le cas d'un SNA $\mathcal{S} = (L, \mathcal{A}, <)$ avec L un langage clos par préfixe et régulier, en commençant par le cas particulier du SNA $\mathcal{S} = (L_k, \mathcal{A}_k, <)$. Avant de passer à ces théorèmes, certains résultats concernant les arbres rationnels doivent encore être établis.

Lemme 3.2.1. [Proposition 1, [4]] *Soit L un langage clos par préfixe et prolongeable sur l'alphabet $(\mathcal{A}, <_{\mathcal{A}})$ et soit $\mathbf{x} = x_0x_1x_2\cdots$ un mot infini sur l'alphabet fini \mathcal{B} . Considérons l'arbre $T(L)$ décoré par le mot \mathbf{x} . L'arbre $T(L)$ est rationnel si et seulement si il existe un entier $h \geq 1$ tel que $\#F_h \leq h$.*

Avant de passer à la preuve de ce lemme, remarquons que pour tout $h \geq 0$ nous avons $\#F_{h+1} \geq \#F_h$. Cela est simplement dû au fait que chaque facteur de hauteur h est un préfixe d'un facteur de hauteur $h + 1$. Le même raisonnement est applicable dans le cas des mots infinis, pour lesquels chaque facteur de longueur $h + 1$ est issu d'un facteur de longueur h . Le nombre de facteurs de longueur $h + 1$ dans un mot infini est donc plus grand ou égal à son nombre de facteurs de longueur h .

Démonstration. Condition nécessaire : Si l'arbre $T(L)$ est rationnel, alors par définition il a un nombre fini k de suffixes. Pour tout entier $h \geq 0$, cela signifie que $\#F_h \leq k$. Nous avons donc la condition de l'énoncé en choisissant $h = k$.

Condition suffisante : Procédons par récurrence sur la hauteur h des facteurs de $T(L)$ et

montrons que si $T(L)$ possède au plus h facteurs de hauteur h alors $T(L)$ est rationnel. Si $h = 1$, tous les noeuds de $T(L)$ ont les mêmes décorations. En effet, en utilisant l'inégalité suivante

$$\#F_0 \leq \#F_1 \leq 1,$$

nous obtenons que le nombre de facteurs de hauteur 0 (i.e., les noeuds de l'arbre) est majoré par 1, donc ils ont tous la même décoration. De plus le fait que $\#F_1 = 1$ permet de conclure que l'arbre est rationnel car la seule façon de construire un facteur de hauteur 2 à partir d'un facteur de hauteur 1 est d'y concaténer l'unique facteur de hauteur 1 comme illustrer à la Figure 3.10.

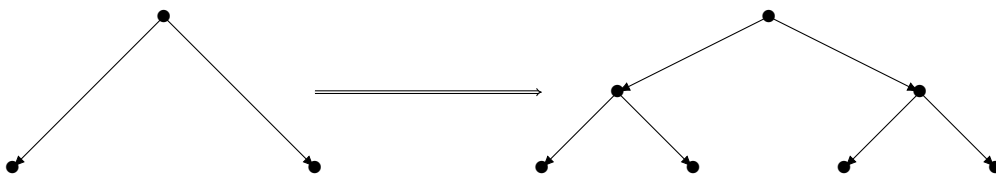


FIGURE 3.10 – Unique facteur de hauteur 1 permettant de construire un facteur de hauteur 2.

Nous avons donc un unique facteurs de hauteur 2 et ainsi de suite.⁵ Supposons maintenant que $h > 1$ et notons $F_{1,h}, \dots, F_{h,h}$ les h facteurs de $T(L)$ de hauteur h . Si on a que $\#F_{h-1} \leq h-1$, alors par hypothèse de récurrence, l'arbre $T(L)$ est rationnel. Dans le cas contraire nous avons que $\#F_{h-1} > h-1$. Tout facteur de hauteur $h-1$ étant un préfixe d'un des sous facteurs $F_{1,h}, \dots, F_{h,h}$ il découle

$$h-1 < \#F_{h-1} \leq \#F_h \leq h. \quad (3.5)$$

Cela implique que $\#F_{h-1} = \#F_h = h$. Nous pouvons alors conclure que l'arbre $T(L)$ est rationnel car cette égalité nous donne $\#F_h = \#F_{h+n} = h$ pour tout $n \geq 0$. Soit $n \geq 0$ et soit $F_{i,h+n}$ un facteur de $T(L)$ de hauteur $h+n$. Ce facteur s'étend en un unique facteur $F_{i,h+n+1}$ de hauteur $h+n+1$ car les suffixes de hauteur $h-1$ de $F_{i,h+n}$ s'étendent de manière unique en facteur de hauteur h au vu de l'égalité (3.5). \square

Au vu de la démonstration du Lemme 3.2.1, nous avons le résultat suivant.

Proposition 3.2.2 ([25], Lemme 26). *Soit L un langage clos par préfixe sur l'alphabet $(\mathcal{A}, <_{\mathcal{A}})$ et soit $\mathbf{x} = x_0x_1x_2 \dots$ un mot infini sur l'alphabet fini \mathcal{B} . Considérons l'arbre $T(L)$ décoré par le mot \mathbf{x} . L'arbre $T(L)$ est rationnel si et seulement si $\#F_h = \#F_{h+1}$ pour un certain $h \geq 0$. En particulier, $\#F_h = \#F_{h+n}$ pour tout $n \geq 0$.*

5. Le fait qu'il ne puisse pas y avoir d'autres facteur de hauteur 2 découle du fait que le langage est prolongeable à droite et donc qu'aucune branche ne peut être finie.

Le théorème suivant caractérise les mots \mathbf{x} k -automatiques en terme de l'arbre $T_{\mathbf{x}}(L_k)$.

Théorème 3.2.3. *Soit $k \geq 2$ un entier. Une suite $\mathbf{x} = x_0x_1x_2\cdots$ est k -automatique si et seulement si l'arbre labellisé $T(L_k)$ décoré par \mathbf{x} est rationnel.*

Démonstration. Condition nécessaire : Si la suite \mathbf{x} est k -automatique alors il existe un AFDS $\mathbb{A} = (Q, q_0, \mathcal{A}_k, \delta, \tau)$ l'engendrant. L'arbre $T(L_k)$ étant un arbre k -aire, tous ses facteurs d'une hauteur fixée le sont aussi. Chaque noeud de l'arbre a k fils, et dans notre cas les arêtes liant chacun des noeuds à ses k fils sont labellisées par $0, 1, \dots, k-1$. En particulier, deux sous-arbres de hauteur fixée ne sont différenciés que par les décorations de leurs noeuds. Soit $w \in L_k$ et un entier $h \geq 0$. Les décorations du facteur $T[w, h]$ de $T(L_k)$ ne dépendent que de la décoration de la racine de $T[w, h]$, qui est donnée par $\tau(\delta(q_0, w))$. Les autres noeuds de $T[w, h]$ sont décorés par $\tau(\delta(q_0, wu))$ pour u parcourant $\mathcal{A}_k^{\leq h}$ dans l'ordre radiciel. Le seul facteur de hauteur h se différenciant des autres de par sa forme est le préfixe $T[\varepsilon, h]$, car la racine n'a que $k-1$ fils (les autres noeuds du préfixe de hauteur h ont k fils). Les arêtes liant la racine à ses fils sont labellisées par $1, \dots, k-1$. Les décorations des noeuds de ce préfixe sont alors données par $\tau(\delta(q_0, u))$ pour u parcourant $\mathcal{A}_k^{\leq h} \setminus 0.\mathcal{A}_k^{< h}$ dans l'ordre radiciel. le nombre de facteur de hauteur h donné par $\#F_h$ est donc majoré par $\#Q + 1$ pour tout $h \geq 0$. La suite $(\#F_h)_{h \geq 0}$ étant croissante et majorée, il existe $H \geq 0$ tel que $\#F_H = \#F_{H+1}$. Nous pouvons dès lors conclure que l'arbre $T(L_k)$ est rationnel en utilisant la Proposition 3.2.2.

Condition suffisante : Supposons que l'arbre $T(L_k)$ est rationnel. Cela signifie qu'il existe $h \geq 0$ tel que $\#F_h = \#F_{h+1}$ par la Proposition 3.2.2. Cette égalité signifie que tout facteur de $T(L_k)$ de hauteur h s'étend de façon unique en un facteur de hauteur $h+1$. Soit f_{h+1} un facteur de $T(L_k)$ de hauteur $h+1$. Dans chacune des copies de f_{h+1} présentes dans l'arbre, la racine a les mêmes k fils de hauteur h . Nous pouvons constater la même chose pour préfixe de hauteur h de $T(L_k)$. Celui-ci s'étend de manière unique en un facteur de hauteur $h+1$, et la racine est liée à $k-1$ sous-arbres de hauteur h ⁶. Afin de montrer l'automaticité de la suite \mathbf{x} , nous allons construire un AFDS \mathcal{F} l'engendrant. Définissons l'ensemble des états de \mathcal{F} comme l'ensemble des facteurs de hauteurs h de $T(L_k)$, noté F_h . La fonction de transition de \mathcal{F} est définie de la façon suivante,

$$\forall i \in \mathcal{A}_k \quad \text{et} \quad \forall w \in \mathcal{A}_k^* : \delta(T[w, h], i) = T[wi, h].$$

L'état initial est donné par $T[\varepsilon, h]$ et posons $\delta(T[\varepsilon, h], 0) = T[\varepsilon, h]$. De cette façon la fonction de transition est bien totale. Le fait que chaque copie d'un même facteur de hauteur $h+1$ de l'arbre $T(L_k)$ ait les mêmes k sous-arbres de hauteur h liés à leur racine

6. Il n'y a que $k-1$ sous-arbres de hauteur h car il n'y a que $k-1$ arêtes sortantes de la racine, qui sont labellisées par $\{1, \dots, k-1\}$.

nous permet de vérifier la condition des AFD donnée à la Remarque 1.2.4. Cela étant, il ne nous reste qu'à définir une fonction de sortie τ pour avoir un AFDS \mathcal{F} . Celle-ci associe à chaque $T[w, h]$ la décoration de sa racine, donnée par $x_{\text{val}_k(w)}$. Vérifions maintenant que \mathcal{F} engendre \mathbf{x} , i.e., pour tout $n \geq 0$ on a $\tau(\delta(T[\varepsilon, h], \text{rep}_k(n))) = x_n$. Par définition du AFDS, nous avons que

$$\tau(\delta(T[\varepsilon, h], \text{rep}_k(n))) = \tau(T[\text{rep}_k(n), h]) = x_{\text{val}_k(\text{rep}_k(n))} = x_n,$$

comme souhaité. □

Ce théorème peut être généralisé au cas de SNA particuliers.

Théorème 3.2.4. *Soit $\mathcal{S} = (L, \mathcal{A}, <)$ un SNA construit sur le langage L clos par préfixe et régulier. Une suite \mathbf{x} est \mathcal{S} -automatique si et seulement si l'arbre $T(L)$ labellisé par \mathbf{x} est rationnel.*

Démonstration. La preuve de ce résultat est presque identique à celle du résultat précédent. Dans la démonstration de celui-ci, les sous-arbres de hauteur h étaient entièrement déterminés par la décoration de leur racine puisqu'ils ont tous les mêmes arêtes (et qui sont labellisées de la même manière). Dans ce cas, un sous-arbre de $T(L)$ noté $T[w, h]$ est à la fois déterminé par $\delta(q_0, w)$ et par $w^{-1}L \cap \mathcal{A}^{\leq h}$. En effet dans la démonstration du Théorème 3.2.3, $w^{-1}L \cap \mathcal{A}^{\leq h}$ n'avait pas d'importance car peu importe le mot w déjà lu dans l'arbre, tout mot u de $\mathcal{A}_k^{\leq h}$ pouvait être lu en partant de la racine de $T[w, h]$. Le langage L étant régulier, l'ensemble $\{w^{-1}L \cap \mathcal{A}^{\leq h} \mid w \in \mathcal{A}^*\}$ est fini. Le nombre de sous-arbres de $T(L)$ de hauteur h est donc majoré par $\#Q$ multiplié par le nombre d'états de l'automate minimal \mathbb{A}_L associé au langage L . □

3.2.1 Caractérisation des suites $\frac{p}{q}$ -automatiques

Les suites $\frac{p}{q}$ -automatiques étant un cas particulier de suites \mathcal{S} -automatiques pour un SNA \mathcal{S} construit sur le langage $L_{\frac{p}{q}}$ qui n'est pas régulier, la caractérisation du Théorème 3.2.4 établie dans la section précédente n'est pas applicable. Notre but est donc d'établir une nouvelle caractérisation utilisant l'arbre $T(L_{\frac{p}{q}})$ destinée aux suites $\frac{p}{q}$ -automatiques.

Soient $m, n \in \mathbb{N}$ deux naturels vérifiant $n < m$. Dans l'arbre $T(L_{\frac{p}{q}})$, le n -ième et le m -ième noeud sont liés par une arête labellisée par $a \in \mathcal{A}_p$ si $\text{rep}_{\frac{p}{q}}(n)a = \text{rep}_{\frac{p}{q}}(m)$ ou de façon équivalente $qm = np + a$. Le résultat suivant nous apprend donc des liens entre deux paires de noeuds de l'arbre $T(L_{\frac{p}{q}})$ dont le premier noeud est lié au deuxième par une arête labellisée par a .

Lemme 3.2.5. *Soient n, n', m, m', p, q et k des entiers. Soit $a \in \{0, \dots, p-1\}$. Supposons que $qm = np + a$ et $qm' = n'p + a$. Dans ce cas,*

- 1) $n \equiv n' \pmod{q^{k+1}}$ si et seulement si $m \equiv m' \pmod{q^k}$,
- 2) $n \equiv n' \pmod{p^k}$ si et seulement si $m \equiv m' \pmod{p^{k+1}}$.

Démonstration. 1) Condition suffisante : Supposons que $m \equiv m' \pmod{q^k}$. Dans ce cas, il existe $b < q^k$ et t, t' des entiers tels que

$$\left. \begin{array}{l} m = tq^k + b \\ m' = t'q^k + b. \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} qm = tq^{k+1} + bq \\ qm' = t'q^{k+1} + bq. \end{array} \right.$$

Nous obtenons donc que $qm \equiv qm' \pmod{q^{k+1}}$. De ce fait,

$$\underbrace{qm - a}_{np} \equiv \underbrace{qm' - a}_{n'p} \pmod{q^{k+1}} \quad \text{i.e.} \quad n \equiv n' \pmod{q^{k+1}},$$

comme souhaité.

Condition nécessaire : Supposons que $n \equiv n' \pmod{q^{k+1}}$. Par un raisonnement similaire à l'autre implication, nous obtenons que

$$qm = qm' \pmod{q^{k+1}} \Rightarrow \left\{ \begin{array}{l} qm = tq^{k+1} + b \\ qm' = t'q^{k+1} + b \end{array} \right.$$

où t, t' sont des entiers. De là nous obtenons

$$\left. \begin{array}{l} m = tq^k + \frac{b}{q} \\ m' = t'q^k + \frac{b}{q} \end{array} \right\} \Rightarrow m \equiv m' \pmod{q^k},$$

car $\frac{b}{q} = m - tq^k$ est entier. Cela conclut. \square

Soit $h \geq$ un entier. La proposition suivante nous apprend que si deux sous-arbres de hauteur h de $T(L_{\frac{p}{q}})$ partagent une racine égale modulo q^h alors ils ont le même domaine et inversement.

Proposition 3.2.6 ([16], Lemme 4.14). *Soient $w, w' \in L_{\frac{p}{q}}$ des mots non vides, et soit h un entier. Alors si $u \in \mathcal{A}_p^h$, nous avons*

- 1) Si $\text{val}_{\frac{p}{q}}(w) \equiv \text{val}_{\frac{p}{q}}(w') \pmod{q^h}$, alors $u \in w^{-1}L_{\frac{p}{q}}$ si et seulement si $u \in (w')^{-1}L_{\frac{p}{q}}$.
- 2) Si $u \in w^{-1}L_{\frac{p}{q}} \cap (w')^{-1}L_{\frac{p}{q}}$, alors $\text{val}_{\frac{p}{q}}(w) \equiv \text{val}_{\frac{p}{q}}(w') \pmod{q^h}$.

Démonstration. 1) Procédons par récurrence sur h . Si $h = 0$, le résultat est vérifié. Supposons que le résultat est vérifié pour tout $h' \leq h$ et montrons que cela reste vrai pour $h+1$. Soit $u \in \mathcal{A}_p^{h+1}$ et montrons que $u \in w^{-1}L_{\frac{p}{q}}$ si et seulement si $u \in (w')^{-1}L_{\frac{p}{q}}$. Par hypothèse, nous avons $\text{val}_{\frac{p}{q}}(w) \equiv \text{val}_{\frac{p}{q}}(w') \pmod{q^h}$, ce qui implique qu'ils sont également congru modulo q (car q^{h+1} est un multiple de q). Notons $u = av$ avec $a \in \mathcal{A}_p$

et $v \in \mathcal{A}_p^h$. Alors $p \cdot \text{val}_{\frac{p}{q}}(w) + a$ est congru à 0 modulo q si et seulement $p \cdot \text{val}_{\frac{p}{q}}(w') + a$ l'est car $\text{val}_{\frac{p}{q}}(w) \equiv \text{val}_{\frac{p}{q}}(w') \pmod{q}$. Supposons qu'ils sont divisibles tous les deux par q , alors il existe des naturels m et m' tels que

$$m = \frac{p \cdot \text{val}_{\frac{p}{q}}(w) + a}{q} \quad \text{et} \quad m' = \frac{p \cdot \text{val}_{\frac{p}{q}}(w') + a}{q}.$$

Nous avons donc les égalités

$$qm = p \cdot \text{val}_{\frac{p}{q}}(w) + a \quad \text{et} \quad qm' = p \cdot \text{val}_{\frac{p}{q}}(w') + a. \quad (3.6)$$

et de façon similaire à la Proposition 2.2.9,

$$m = \text{val}_{\frac{p}{q}}(wa) \quad \text{et} \quad m' = \text{val}_{\frac{p}{q}}(w'a).$$

Il découle donc que wa et $w'a$ sont des mots de $L_{\frac{p}{q}}$ car ils s'écrivent comme la représentation en base $\frac{p}{q}$ d'un nombre naturel. En utilisant l'Égalité 3.6, le Lemme 3.2.5 et l'hypothèse nous obtenons que $m \equiv m' \pmod{q^h}$ et donc $\text{val}_{\frac{p}{q}}(wa) \equiv \text{val}_{\frac{p}{q}}(w'a) \pmod{q^h}$. Nous pouvons donc appliquer l'hypothèse de récurrence aux deux mots $wa, w'a \in L_{\frac{p}{q}}$ et au mot $v \in \mathcal{A}_p^h$. Cela implique de $v \in (wa)^{-1}L_{\frac{p}{q}}$ si et seulement si $v \in (w'a)^{-1}L_{\frac{p}{q}}$ et donc $av \in (w)^{-1}L_{\frac{p}{q}}$ si et seulement si $av \in (w')^{-1}L_{\frac{p}{q}}$.

2) Procédons également par récurrence sur h . Le cas $h = 0$ étant trivialement vérifié, supposons que le proposition est vraie pour tout $h' \leq h$ et montrons qu'il reste vrai pour $h+1$. Soit $u \in \mathcal{A}_p^{h+1}$ un mot de longueur $h+1$ tel que $u \in w^{-1}L_{\frac{p}{q}} \cap (w')^{-1}L_{\frac{p}{q}}$. Notons $u = av$ avec $a \in \mathcal{A}_p$ et $v \in \mathcal{A}_p^h$. Puisque le langage $L_{\frac{p}{q}}$ est clos par préfixe, le fait que $wu \in L_{\frac{p}{q}}$ implique que $wa \in L_{\frac{p}{q}}$. Par hypothèse, les mots wu et $w'u$ étant des mots de $L_{\frac{p}{q}}$, nous avons $v \in (wa)^{-1}L_{\frac{p}{q}} \cap (w'a)^{-1}L_{\frac{p}{q}}$. Par hypothèse de récurrence, cela implique que $\text{val}_{\frac{p}{q}}(wa) \equiv \text{val}_{\frac{p}{q}}(w'a) \pmod{q^h}$.

Par définition de $L_{\frac{p}{q}}$, il existe des naturels m et m' tels que $\text{rep}_{\frac{p}{q}}(m) = wa$ et $\text{rep}_{\frac{p}{q}}(m') = w'a$. Nous avons donc que par la Propriété (2.2),

$$m = \text{val}_{\frac{p}{q}}(wa) = \text{val}_{\frac{p}{q}}(w) \cdot \frac{p}{q} + \frac{a}{q} \quad \text{et} \quad m' = \text{val}_{\frac{p}{q}}(w'a) = \text{val}_{\frac{p}{q}}(w') \cdot \frac{p}{q} + \frac{a}{q}.$$

Par le Lemme 3.2.5, nous obtenons que $\text{val}_{\frac{p}{q}}(w) \equiv \text{val}_{\frac{p}{q}}(w') \pmod{q^{h+1}}$, comme souhaité. \square

Proposition 3.2.7 (Lemme 4.15, [16]). *Soient $w, w' \in L_{\frac{p}{q}}$ des mots non vides et soit h un naturel. Alors si $u \in \mathcal{A}_p^h$, nous avons*

- 1) Si $\text{val}_{\frac{p}{q}}(w) \equiv \text{val}_{\frac{p}{q}}(w') \pmod{p^h}$, alors $u \in \text{Suff}(w)$ si et seulement si $u \in \text{Suff}(w')$,
- 2) Si $u \in \text{Suff}(w) \cap \text{Suff}(w')$, alors $\text{val}_{\frac{p}{q}}(w) \equiv \text{val}_{\frac{p}{q}}(w') \pmod{p^h}$.

Démonstration. 1) Procédons par récurrence sur h . Si $h = 0$, le résultat est trivialement vérifié. Supposons que le résultat soit vérifié pour tout $h' \leq h$ et montrons pour $h + 1$. Soit $u = va$ un mot de longueur $h + 1$ tel que $v \in \mathcal{A}_p^h$ et $a \in \mathcal{A}_p$. Montrons que $a \in \text{Suff}(w)$ si et seulement si $a \in \text{Suff}(w')$. Soit $t', t \in \mathcal{A}_p^*$ et $b \in \mathcal{A}_p$ tel que $w = ta$ et $w' = t'b$. Cela implique que

$$q \cdot \text{val}_{\frac{p}{q}}(w) = \text{val}_{\frac{p}{q}}(t) \cdot p + a \quad \text{et} \quad q \cdot \text{val}_{\frac{p}{q}}(w') = \text{val}_{\frac{p}{q}}(t') \cdot p + b. \quad (3.7)$$

Par hypothèse nous savons que $\text{val}_{\frac{p}{q}}(w) \equiv \text{val}_{\frac{p}{q}}(w') \pmod{p^{h+1}}$, ce qui implique que nous avons également $\text{val}_{\frac{p}{q}}(w) \equiv \text{val}_{\frac{p}{q}}(w') \pmod{p}$. Grâce à cela et au vu de l'égalité (3.7), nous avons

$$\left. \begin{array}{l} q \cdot \text{val}_{\frac{p}{q}}(w) \equiv a \pmod{p} \\ q \cdot \text{val}_{\frac{p}{q}}(w') \equiv b \pmod{p} \end{array} \right\} \Rightarrow q \cdot \text{val}_{\frac{p}{q}}(w) \equiv q \cdot \text{val}_{\frac{p}{q}}(w') \pmod{p} \Rightarrow a = b,$$

et les mots w et w' finissent par la même lettre. De là, par le Lemme 3.2.5, nous avons $\text{val}_{\frac{p}{q}}(t) \equiv \text{val}_{\frac{p}{q}}(t') \pmod{p^h}$. Les mots t et t' étant des préfixes de mots de $L_{\frac{p}{q}}$, ils sont également des mots de $L_{\frac{p}{q}}$ de longueur h auxquels nous pouvons appliquer l'hypothèse de récurrence. De ce fait, $v \in \text{Suff}(t)$ si et seulement si $v \in \text{Suff}(t')$, ce qui implique que $va \in \text{Suff}(w)$ si et seulement si $va \in \text{Suff}(w')$ comme souhaité.

2) Procédons par récurrence sur h . Si $h = 0$, le résultat est vrai. Supposons que le résultat soit vérifié pour tout $h' \leq h$, et montrons qu'il reste vérifié pour $h + 1$. Considérons $u \in \mathcal{A}_p^{h+1}$, et notons le $u = va$ où $a \in \mathcal{A}_p$ et $v \in \mathcal{A}_p^h$. Le mot u étant par hypothèse un suffixe des mots w et w' et le langage $L_{\frac{p}{q}}$ étant clos par préfixe, il existe des mots $n, n' \in L_{\frac{p}{q}}$ tels que $w = na$ et $w' = n'a$. Nous pouvons donc appliquer l'hypothèse de récurrence aux mots n et n' , et au mot v vérifiant $v \in \text{Suff}(n) \cap \text{Suff}(n')$. Nous avons donc $\text{val}_{\frac{p}{q}}(n) \equiv \text{val}_{\frac{p}{q}}(n') \pmod{p^h}$, ce qui permet de conclure grâce au Lemme 3.2.5 en remarquant que

$$\text{val}_{\frac{p}{q}}(w) = \text{val}_{\frac{p}{q}}(na) \Rightarrow q \cdot \text{val}_{\frac{p}{q}}(w) = \text{val}_{\frac{p}{q}}(n) \cdot p + a$$

et

$$\text{val}_{\frac{p}{q}}(w') = \text{val}_{\frac{p}{q}}(n'a) \Rightarrow q \cdot \text{val}_{\frac{p}{q}}(w') = \text{val}_{\frac{p}{q}}(n') \cdot p + a$$

□

Corollaire 3.2.8. *Tout mot $u \in \mathcal{A}_p^*$ est suffixe d'un mot de $L_{\frac{p}{q}}$.*

Démonstration. Soit h un naturel. Par le Lemme 3.2.7, nous savons que deux noeuds n, m sont atteints par le même mot u de longueur h si et seulement les mots $w, w' \in L_{\frac{p}{q}}^p$ lus dans l'arbre depuis la racine pour atteindre les noeuds n et m vérifient $\text{val}_{\frac{p}{q}}(w) \equiv \text{val}_{\frac{p}{q}}(w') \pmod{p^h}$. Pour tous les mots $w \in L_{\frac{p}{q}}^p$ liant la racine à un noeud de niveau plus grand ou égal à h de l'arbre, il existe $R \in \{0, \dots, p^h - 1\}$ tel que $\text{val}_{\frac{p}{q}}(w) \equiv R \pmod{p^h}$. De plus, si on note $n_0, n_1, \dots, n_{p^h-1}$ les noeuds consécutifs de niveau plus grand ou égal à h liés à la racine par les mots $w_0, w_1, \dots, w_{p^h-1} \in L_{\frac{p}{q}}^p$ et tel que $\text{val}_{\frac{p}{q}}(w_0) \equiv 0 \pmod{p^h - 1}$, alors $\text{val}_{\frac{p}{q}}(w_i) \equiv i \pmod{p^h - 1}$ pour tout $i \in \{1, \dots, p^h - 1\}$. Chacun de ces mots appartient donc à une classe de congruence différente et ils ont donc tous des suffixes distincts de longueur h . Le nombre de mots de \mathcal{A}_p^h étant p^h , ils sont chacun d'entre eux suffixe d'un des w_i et tous les mots de \mathcal{A}_p de longueur h sont bien suffixes d'un mot de $L_{\frac{p}{q}}^p$. \square

Soit $u \in \mathcal{A}_p^h$. Le Corollaire 3.2.8 nous apprend qu'il existe $w, v \in L_{\frac{p}{q}}^p$ tels que $v = wu$ et donc vérifiant $u \in w^{-1}L_{\frac{p}{q}}^p \cap \mathcal{A}_p^h$. Le nombre de langages de la forme $w^{-1}L_{\frac{p}{q}}^p \cap \mathcal{A}_p^h$ pour $w \in L_{\frac{p}{q}}^p \setminus \{\varepsilon\}$ étant au nombre de q^h par le Lemme 3.2.6, la langage \mathcal{A}_p^h est partitionné en q^h langages donnés par $\{w^{-1}L_{\frac{p}{q}}^p \cap \mathcal{A}_p^h \mid w \in L_{\frac{p}{q}}^p \setminus \{\varepsilon\}\}$. Cela implique donc que pour tout $h \geq 0$ entier, le nombre $\#F_h$ de facteurs de hauteur h de l'arbre $T(L_{\frac{p}{q}}^p)$ vérifie $\#F_h \leq q^h + 1$ ⁷

Exemple 3.2.9. Nous pouvons illustrer comment les facteurs de hauteur 1 s'étendent en facteurs de hauteur 2 dans l'arbre $T(L_{\frac{3}{2}})$ non labellisé. Au vu de la signature périodique labellisée $\mathbf{s} = (02, 1)^\omega$ associée à cet arbre, la seule possibilité est d'étendre une feuille par une ou deux arêtes, labellisées par les mots de \mathbf{s} . De plus, si on étend l'arbre par deux arêtes sur la n -ième feuille, la $(n+1)$ -ième feuille doit être étendue par une unique arête pour respecter la forme de signature labellisée \mathbf{s} . De cette façon, deux feuilles successives ne peuvent pas être étendues par deux arêtes chacune. Le facteur de longueur 1 de la racine est traité différemment et ne peut être étendu que d'une seule manière en un facteur de hauteur 2. Une représentation des facteurs de hauteur 1 et 2 de $T(L_{\frac{3}{2}})$ est donnée à la Figure 3.11. Nous pouvons remarquer que le nombre de facteurs de hauteur 1 vérifie $\#F_1 = 3 \leq 2^1 + 1$. C'est également le cas pour les facteurs de hauteur 2, $\#F_2 = 5 \leq 2^2 + 1$.

Soit $\mathbf{x} \in \mathcal{B}^{\mathbb{N}}$ où \mathcal{B} est un alphabet fini. Nous allons maintenant établir un résultat similaire en décorant l'arbre $T(L_{\frac{p}{q}}^p)$ grâce au mot \mathbf{x} .

7. Le Lemme 3.2.6 traite le cas des sous-arbres dont la racine n'est pas ε (i.e. les sous-arbres qui ne sont pas le préfixe de hauteur h). Celui-ci est donc rajouté dans la borne concernant le nombre total de facteurs de hauteur h .

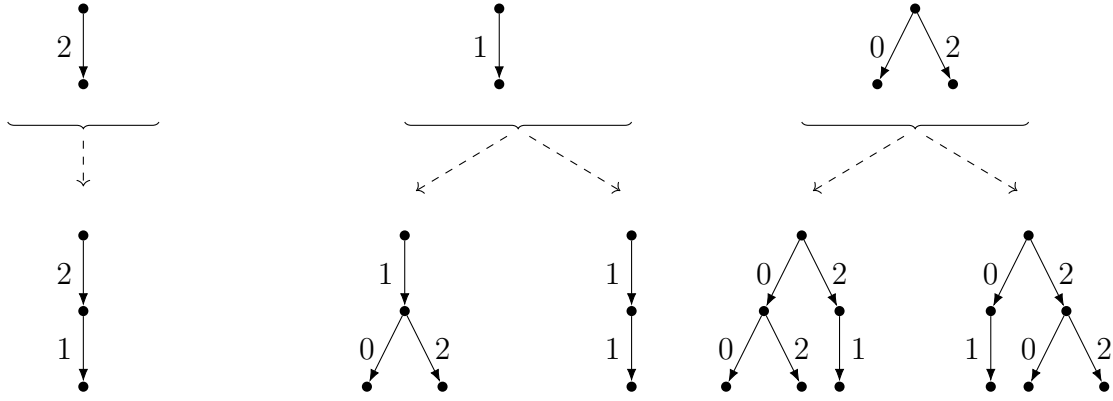


FIGURE 3.11 – Extensions des facteurs de hauteur 1 en facteurs de hauteur 2 dans l'arbre $T(L_{\frac{3}{2}})$.

Lemme 3.2.10. Soit \mathbf{x} une suite $\frac{p}{q}$ -automatique produite par le AFDS $\mathbb{A} = (Q, q_0, \mathcal{A}_p, \delta, \tau)$ et soit $T(L_{\frac{p}{q}})$ l'arbre décoré par \mathbf{x} . Pour tout $h \geq 1$, le nombre $\#F_h$ de facteurs de hauteur h de $T(L_{\frac{p}{q}})$ est majoré par $q^h \cdot \#Q + 1$.

Démonstration. Soit $h \geq 1$ et soient $w, w' \in L_{\frac{p}{q}}$ un mot non vide. Le facteur de hauteur h noté $T[w, h]$ est entièrement déterminé par w . Pour comparer des sous-arbres de $T(L_{\frac{p}{q}})$ de hauteur h , il nous suffit de connaître leur domaine et leur décoration. Par le Lemme 3.2.6, deux sous-arbres $T[w, h]$ et $T[w', h]$ de hauteur h ont le même domaine si $\text{val}_{\frac{p}{q}}(w) \equiv \text{val}_{\frac{p}{q}}(w') \pmod{q^h}$. Nous avons donc au plus q^h sous-arbres de hauteur h ayant un domaine distinct. Pour un sous-arbre de domaine fixé, les décorations du sous-arbre dépendent uniquement de $\delta(q_0, w)$ et sont données par $\tau(\delta(q_0, wu))$ pour u parcourant $w^{-1}L_{\frac{p}{q}} \cap \mathcal{A}_p^{\leq h}$. Le nombre de sous-arbres distincts pour un domaine fixé est donc donné par le nombre d'état de \mathbb{A} . Cela implique donc que le nombre de sous-arbres de hauteur h de la forme $T[w, h]$ est majoré par $Q \cdot q^h$. À cela s'ajoute le préfixe $T[\varepsilon, h]$ de l'arbre, dont les décorations sont données par $\tau(\delta(q_0, u))$ pour u variant dans $L_{\frac{p}{q}} \cap \mathcal{A}_p^{\leq h}$. Nous avons donc finalement $\#F_h \leq Q \cdot q^h + 1$. \square

Le Lemme 3.2.10 nous permet donc d'établir la caractérisation des suites $\frac{p}{q}$ -automatiques en terme du nombre de facteurs de l'arbre $T(L_{\frac{p}{q}})$. Cette caractérisation se décompose en la Proposition 3.2.11 et la Proposition 3.2.13.

Proposition 3.2.11. Soit \mathbf{x} une suite $\frac{p}{q}$ -automatique sur un alphabet fini \mathcal{B} , générée par le AFDS $\mathbb{A} = (Q, q_0, \mathcal{A}_p, \delta, \tau, \mathcal{B})$ ayant la propriété suivante : il existe un entier h tel que pour tous mots $u, v \in L_{\frac{p}{q}}$ tels que $\text{val}_{\frac{p}{q}}(u) \equiv \text{val}_{\frac{p}{q}}(v) \pmod{q^h}$ et $\delta(q_0, u) \neq \delta(q_0, v)$, il existe un mot $w \in u^{-1}L_{\frac{p}{q}} \cap \mathcal{A}_p^{\leq h} = v^{-1}L_{\frac{p}{q}} \cap \mathcal{A}_p^{\leq h}$ tel que $\tau(\delta(q_0, uw)) \neq \tau(\delta(q_0, vw))$.

8. Cette égalité est due à la Proposition 3.2.6 car les mots u et v vérifient $\text{val}_{\frac{p}{q}}(u) \equiv \text{val}_{\frac{p}{q}}(v) \pmod{q^h}$

Alors, pour tout $0 \leq j \leq q - 1$ nous avons dans l'arbre $T(L_{\frac{p}{q}})$ décoré par \mathbf{x} ,

$$\#F_{h+1, w_j, 0}^{\infty} \leq F_h^{\infty}.$$

Remarque 3.2.12. La condition concernant l'AFDS \mathbb{A} de l'énoncé peut être atteinte en appliquant un algorithme de minimisation à l'automate. Cet algorithme permet de déterminer si deux états q, q' sont distingués, i.e. s'il existe un mot $u \in \mathcal{A}_p^*$ tel que $\tau(\delta(q, u)) \neq \tau(\delta(q', u))$. Pour ce faire, toutes les paires d'états $\{q, q'\}$ sont passées en revue pour vérifier si elles sont distinguées par le mot vide, i.e. si $\tau(q) \neq \tau(q')$. De là, nous passons en revue toutes les paires d'états $\{q, q'\}$ n'ayant pas encore été distingués. Si pour une paire d'états $\{q, q'\}$ pas encore distingués il existe $a \in \mathcal{A}_p$ tel que $\tau(\delta(q, a)) = p$ et $\tau(\delta(q', a)) = p'$ (ou inversement) pour $\{p, p'\}$ une paire d'états distingués, alors la paire d'états $\{q, q'\}$ est également distinguée. L'algorithme s'arrête lorsqu'aucune paire d'états supplémentaire ne peut être distingués. Nous pouvons dès lors définir un nouvel AFDS dans lequel les états non distingués par l'algorithme peuvent alors être rassemblés en un seul état.

La condition supplémentaire $w \in u^{-1}L_{\frac{p}{q}} \cap \mathcal{A}_p^{\leq h} = v^{-1}L_{\frac{p}{q}} \cap \mathcal{A}_p^{\leq h}$ sur le mot w qui distingue la paire d'états $\{q, q'\}$ est là pour assurer que le mot qui distingue deux états est lisible depuis les deux états en question. En effet, l'AFDS \mathbb{A} de l'énoncé permet la lecture de mots de $L_{\frac{p}{q}}$, qui ne sont qu'une sous partie des mots de \mathcal{A}_p^* . Pour les mots k -automatiques, cette condition n'a pas lieu d'être car tout mot $w \in \mathcal{A}_k^*$ peut être lu depuis n'importe quel état.

Démonstration. Commençons par remarquer que les facteurs de hauteur $h + 1$ apparaissant infiniment souvent dans $T(L_{\frac{p}{q}})$ ne peuvent être que des extensions d'éléments de F_h^{∞} . Nous allons donc regarder le nombre d'extension maximal pouvant être produit par un élément de F_h^{∞} .

Soit donc un facteur de hauteur h de $T(L_{\frac{p}{q}})$ apparaissant infiniment souvent, i.e. il existe une suite $(u_i)_{i \geq 1}$ de mots de $L_{\frac{p}{q}}$ tels que $T[u_1, h] = T[u_2, h] = \dots$. La condition sur le AFDS de l'énoncé signifie que si les états q, q' atteints en lisant deux mots $w, v \in L_{\frac{p}{q}}$ vérifiant $w^{-1}L_{\frac{p}{q}} \cap \mathcal{A}_p^{\leq h} = v^{-1}L_{\frac{p}{q}} \cap \mathcal{A}_p^{\leq h}$ sont distincts, alors il existe $u \in w^{-1}L_{\frac{p}{q}} \cap \mathcal{A}_p^{\leq h} = v^{-1}L_{\frac{p}{q}} \cap \mathcal{A}_p^{\leq h}$ tel que $\tau(\delta(q, u)) \neq \tau(\delta(q', u))$. En particulier, si $\delta(q_0, w) \neq \delta(q_0, v)$ alors $T[w, h] \neq T[v, h]$. Par contraposition nous avons que pour tout $i, j \geq 1$, $T[u_i, h] = T[u_j, h]$ implique $\delta(q_0, u_i) = \delta(q_0, u_j)$. Ainsi, si $T[u_i, h + 1]$ et $T[u_j, h + 1]$ ont le même domaine, alors ils sont égaux car $\delta(q_0, u_i v) = \delta(q_0, u_j v)$ pour tout $v \in \text{dom}(T[u_i, h + 1]) = \text{dom}(T[u_j, h + 1])$.

Deux sous-arbres distincts de hauteur $h + 1$ ayant le même domaine ne peuvent donc être issus de l'extension d'un même sous-arbre de hauteur h (si ils provenaient du même sous-arbre et qu'ils avaient le même domaine, ils seraient égaux par le raisonnement précédent). Un facteur de hauteur h ne peut donc s'étendre qu'en au plus un

sous-arbre de hauteur $h + 1$ ayant $w_{j,0}$ labellisant la première arête joignant le niveau h au niveau $h + 1$, pour $0 \leq j \leq q - 1$. Cela implique donc l'inégalité de l'énoncé. \square

Proposition 3.2.13. *Soit \mathbf{x} une suite sur un alphabet fini \mathcal{B} et soit $T(L_{\frac{p}{q}})$ l'arbre décoré par \mathbf{x} . Si il existe un $h \geq 0$ tel que pour tout $0 \leq j \leq q - 1$ nous avons $\#F_{h+1, w_{j,0}}^{\infty} \leq F_h^{\infty}$, alors \mathbf{x} est $\frac{p}{q}$ -automatique.*

Démonstration. Pour montrer que \mathbf{x} est $\frac{p}{q}$ -automatique, nous allons construire un AFND qui sera ensuite transformer en un AFDS engendrant le mot \mathbf{x} .

Définissons un AFND \mathbb{A} de la façon suivante. Les préfixes de hauteurs h n'apparaissant qu'un nombre fini de fois se trouvent dans le préfixe de $T(L_{\frac{p}{q}})$ (que nous notons T). Soit donc t le plus petit entier tel que les facteurs de hauteur h ayant pour racine un noeud à un niveau l vérifiant $l \geq t$ soient des éléments de F_h^{∞} . Le préfixe $T[\varepsilon, t + h - 1]$ contient donc tous les facteurs de hauteur h n'apparaissant qu'un nombre fini de fois dans T . L'automate \mathbb{A} est composé des noeuds et transitions du préfixe $T[\varepsilon, t + h - 1]$ ainsi que des noeuds et transitions de chaque élément de F_h^{∞} . Il reste donc à définir les transitions entre les éléments de F_h^{∞} et le préfixe $T[\varepsilon, t + h - 1]$.

a) Transitions entre les noeuds de $T[\varepsilon, t + h - 1]$ et les éléments de F_h^{∞} .

Si un noeud m au niveau $t - 1$ du préfixe $T[\varepsilon, t + h - 1]$ est lié à un noeud n au niveau t par une arête labellisée par d , il y aura une transition allant du noeud m de $T[\varepsilon, t + h - 1]$ vers la racine de l'arbre $T[n, h] \in F_h^{\infty}$. L'arbre $T[n, h]$ est un élément de F_h^{∞} car le noeud n est au niveau t .

b) Transitions entre les éléments de F_h^{∞} .

Soit r la racine d'un arbre $T[r, h] \in F_h^{\infty}$. Supposons que r à un fils, noté s , auquel il est relié par un arête labellisée par d . L'hypothèse de l'énoncé implique que $T[r, h]$ s'étend en au plus un élément U_c de $F_{h+1, c}^{\infty}$ pour $c \in \{w_{0,0}, w_{1,0}, \dots, w_{q-1,0}\}$. L'arbre U_c a donc un sous-arbre $V_{c,d}$ de hauteur h et de racine s . Pour un d fixé, il y a au plus q arbres $V_{c,d} \in F_h^{\infty}$ pour c variant dans $\{w_{0,0}, w_{1,0}, \dots, w_{q-1,0}\}$. Dans l'automate \mathbb{A} , il y aura une transition entre le noeud r (qui est la racine de $T[r, h] \in F_h^{\infty}$) et le noeud s (qui est la racine de $V_{c,d} \in F_h^{\infty}$) labellisée par la lettre d .

Le seul état initial de \mathbb{A} est la racine du préfixe $T[\varepsilon, t + h - 1]$. Ses états finaux sont l'ensemble des noeuds de $T[\varepsilon, t + h - 1]$ et les noeuds de niveau h pour chaque élément de F_h^{∞} .

Nous allons maintenant montrer que tout mot $u \in L_{\frac{p}{q}}$ est accepté par \mathbb{A} et qu'il n'existe qu'un unique trajet dans l'automate pour l'accepter. Soit donc $u \in L_{\frac{p}{q}}$.

- Si $|u| \leq t + h$.

Il n'y a alors qu'un seul trajet possible dans \mathbb{A} , qui se fait entièrement dans le préfixe $T[\varepsilon, t + h - 1]$. Chaque noeud de ce préfixe étant final, le mot u est accepté par \mathbb{A} dans ce cas.

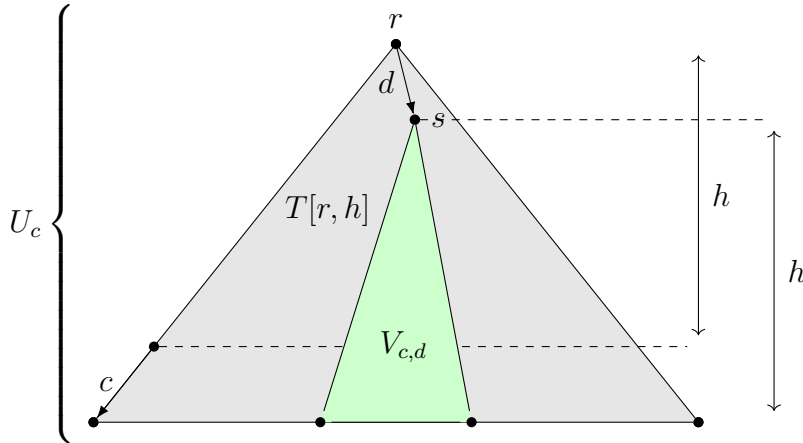


FIGURE 3.12 – Transitions entre deux arbres de F_h^∞ .

- Si le trajet dans \mathbb{A} emprunte une transition définie au point a .
Si le trajet emprunté dans \mathbb{A} utilise une transition définie au point a , le mot u doit nécessairement être de longueur supérieure ou égale à $t + h$. En effet, les transitions définies au point a mènent à la racine d'un élément de F_h^∞ qui ont pour seuls états finaux leurs noeuds de niveau h .
Soit $j \geq 0$ tel que u s'écrit de la façon suivante

$$u = u_0 \cdots u_{t-1} u_t u_{t+1} \cdots u_{t+h-1} \cdots u_{t+h+j-1}.$$

La lecture du préfixe $u_0 \cdots u_{t-1}$ mène dans l'automate à la racine d'un élément U de F_h^∞ . Supposons que ce facteur de hauteur h s'étende en au moins deux facteurs de hauteur $h + 1$, appartenant respectivement à F_{h+1, c_1}^∞ et F_{h+1, c_2}^∞ . Cela signifie, en utilisant les transitions définies au point b , que dans \mathbb{A} il y a deux transitions partant de la racine (respectivement notée t_1 et t_2) menant à deux sous-arbres $V_1 \in F_{h, c_1}^\infty$ et $V_2 \in F_{h, c_2}^\infty$ vérifiant $V_1[\varepsilon, h - 1] = V_2[\varepsilon, h - 1]$. Les deux sous-arbres V_1 et V_2 sont uniquement différenciés par le choix de c_1, c_2 parmi $\{w_{0,0}, w_{1,0}, \dots, w_{q-1,0}\}$. Si $c_1 = w_{j,0}$ les labels des arêtes liant le niveau $h - 1$ au niveau h de V_1 sont donnés par la signature périodique labellisée $(w_j, \dots, w_{q-1}, w_0, \dots, w_{j-1})$. Les deux extensions de U sont représentées à la Figure 3.13. De cette façon, si un mot x de longueur $h - 1$ appartient au domaine de V_1 (et donc à celui de V_2), alors xc_1 appartient au domaine de V_1 si et seulement si xc_2 appartient au domaine de V_2 . Au moment de la lecture de u_t dans l'automate \mathcal{A} , si la mauvaise transition est choisie de façon non déterministe, la lettre u_{t+h} ne pourra pas être lue. Nous pouvons raisonner de la même façon au moment de la lecture de la lettre u_{t+j} , pour être capable de lire la lettre u_{t+j+h} . Le chemin menant à l'acceptation du mot u est donc unique.⁹

9. A noter que le mot u est accepté par construction de l'arbre T .

Nous pouvons maintenant transformer l'automate en AFDS \mathbb{A} produisant $\mathbf{x} \in \mathcal{B}^{\mathbb{N}}$, i.e. un AFD muni d'une fonction de sortie $\tau : \mathcal{A}_p \rightarrow \mathcal{B}$. L'AFND peut être transformé en un premier temps en AFD grâce à la construction par sous-ensembles. Comme nous l'avons montré, tout mot $u \in L_{\frac{p}{q}}$ est accepté par l'automate \mathbb{A} et il n'existe qu'un parcours dans l'automate menant à cette acceptation. De ce fait, dans l'automate déterminisé obtenu à partir de \mathbb{A} , un état (qui est un sous ensemble d'états de l'automate \mathbb{A}) ne peut contenir qu'un état final au plus. Si un état q de l'automate déterminisé contient un état final, alors nous posons $\tau(q) = b$ où b est la décoration de l'état final faisant partie du sous ensemble d'états q . Si il n'en contient pas, la sortie correspondante à cet état n'a pas d'importance et peut être choisie au hasard.

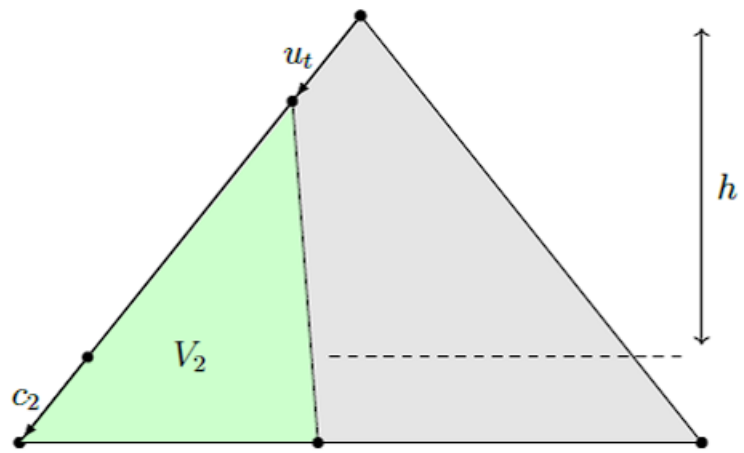
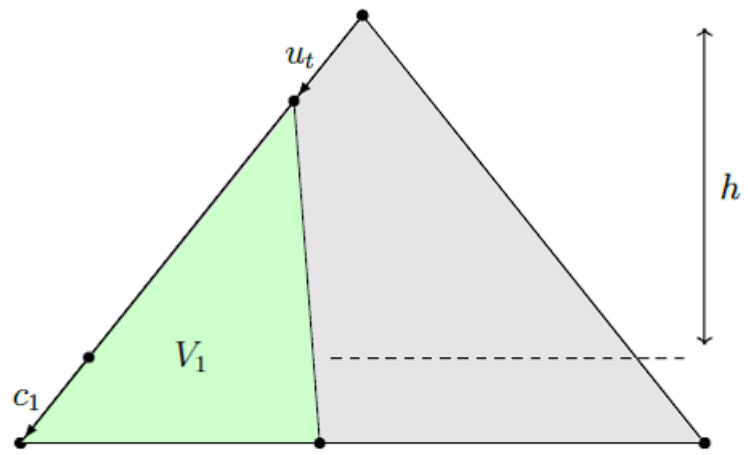


FIGURE 3.13 – Les deux extensions de l'arbre U .

□

Chapitre 4

Suites $\frac{p}{q}$ -régulières

Ce chapitre est consacré à la généralisation des suites k -régulières à un SNA \mathcal{S} construit sur un langage régulier afin de définir les suites \mathcal{S} -régulières. Pour ce faire, nous commençons par rappeler deux définitions des suites k -régulières : la définition en terme de son k -noyau et en terme de la série formelle particulière qui lui est associée. Ensuite, nous généralisons ces définitions aux SNA construits sur un langage régulier en introduisant la notion de \mathcal{S} -noyau. Nous caractériserons ensuite les suites \mathcal{S} -régulières grâce à un nouveau type d'arbre appelé arbre h -linéaire. Pour finir, nous établissons une définition des suites $\frac{p}{q}$ -régulières en se servant de la caractérisation établie concernant les suites \mathcal{S} -régulières pour un SNA construit sur un langage L régulier. La définition des suites $\frac{p}{q}$ -régulières se servant de la h -linéarité de l'arbre $T(L_{\frac{p}{q}})$ nous décrivons dans la dernière section une procédure permettant d'essayer d'établir la h -linéarité de celui-ci.

4.1 Suites k -régulières

Dans le Chapitre 3, nous avons introduit la notion de mots k -automatiques et $\frac{p}{q}$ -automatiques. La généralisation des mots k -automatiques à un alphabet infini \mathbb{K} a pour la première fois été introduite dans l'article [3] et porte le nom de mots k -réguliers. Considérons \mathbb{L} un anneau Noethérien et supposons que \mathbb{L} est un sous semi-anneau de \mathbb{K} . Pour rappel, un mot $\mathbf{x} \in \mathbb{K}^{\mathbb{N}}$ est (\mathbb{L}, k) -régulier si le \mathbb{L} -module engendré par les sous-suites de son k -noyau, noté $\ker_k(\mathbf{x})$, et défini comme

$$\ker_k(\mathbf{x}) = \{(x_{k^\ell n+r})_{n \geq 0}, \mid \ell \geq 0, 0 \leq r < k^\ell\} \quad (4.1)$$

est finiment engendré. Le \mathbb{L} -module engendré par les sous-suites de $\ker_k(\mathbf{x})$ est noté $\langle \ker_k(\mathbf{x}) \rangle_{\mathbb{L}}$. Le \mathbb{L} -module engendré par un ensemble B noté $\langle B \rangle_{\mathbb{L}}$ est l'ensemble des combinaisons linéaires de la forme $l_1 b_1 + l_2 b_2 + \dots + l_n b_n$ telles que $l_i \in \mathbb{L}$ et $b_i \in B$

pour tout $i \in \{1, \dots, l\}$.

Remarque 4.1.1. Nous choisissons l'anneau \mathbb{L} Noethérien car la définition des suites k -régulières se sert du \mathbb{L} -module engendré par le k -noyau. En effet, si $\mathbf{x} \in \mathbb{K}^{\mathbb{N}}$ et si \mathbb{L} est Noethérien, alors tout sous-module de $\langle \ker(\mathbf{x}) \rangle_{\mathbb{L}}$ est finiment engendré si $\langle \ker(\mathbf{x}) \rangle_{\mathbb{L}}$ est lui-même finiment engendré [11]. Ce résultat nous sera par exemple utile dans le Lemma 4.2.13.

Exemple 4.1.2. Considérons la suite \mathbf{s}_2 définie pour tout $n \geq 0$ entier comme $\mathbf{s}_2(n) = |\text{rep}_2(n)|_1$, i.e. la suite de la somme des chiffres de la représentation en base 2. Cette suite est 2-régulière. En effet, pour $\ell = 1$ et $r \in \{0, 1\}$ dans la définition du 2-noyau, nous avons les sous-suites suivantes

$$(\mathbf{s}_2(2n))_{n \geq 0} = \mathbf{s}_2(0)\mathbf{s}_2(2)\mathbf{s}_2(4) \cdots \text{ et } (\mathbf{s}_2(2n+1))_{n \geq 0} = \mathbf{s}_2(1)\mathbf{s}_2(3)\mathbf{s}_2(5) \cdots .$$

La multiplication par 2 d'un naturel ajoutant 0 à sa représentation en base 2, nous savons donc que pour tout $n \geq 0$, $\mathbf{s}_2(2n) = \mathbf{s}_2(n)$. De façon similaire, $\mathbf{s}_2(2n+1) = \mathbf{s}_2(n) + (1)_{n \geq 0}$. Les premières suites du 2-noyau peuvent donc être réexprimées comme une combinaison linéaire des suites \mathbf{s}_2 et $(1)_{n \geq 0}$. Les autres éléments du 2-kernel étant des sous-suites des suites $\mathbf{s}_2(2n)$ et $\mathbf{s}_2(2n+1)$ qui sont réexprimables en terme de \mathbf{s}_2 et $(1)_{n \geq 0}$, nous obtenons

$$\langle \ker_2(\mathbf{s}_2) \rangle_{\mathbb{N}} = \langle \mathbf{s}_2, (1)_{n \geq 0} \rangle_{\mathbb{N}}.$$

La suite \mathbf{s}_2 est donc $(\mathbb{N}, 2)$ -régulière.

Remarque 4.1.3 (Théorème 2.2,[3]). Si B est un ensemble tel que $\langle B \rangle_{\mathbb{L}}$ est finiment engendré, il existe des éléments s_1, \dots, s_m tels que

$$\langle B \rangle_{\mathbb{L}} = \langle s_1, \dots, s_m \rangle_{\mathbb{L}}.$$

Bien que s_1, \dots, s_m ne soient pas nécessairement des éléments de B , au vu de l'égalité entre les deux \mathbb{L} -modules engendrés nous avons

$$\langle s_1, \dots, s_m \rangle_{\mathbb{L}} \subset \langle B \rangle_{\mathbb{L}}.$$

En particulier, il existe pour tout $i \in \{1, \dots, m\}$ des éléments $l_{i_1}, l_{i_2}, \dots, l_{i_{j_i}} \in \mathbb{L}$ et des éléments $b_{i_1}, \dots, b_{i_{j_i}} \in B$ tels que

$$s_i = l_{i_1} b_{i_1} + \cdots + l_{i_{j_i}} b_{i_{j_i}}.$$

Les éléments s_i sont donc pour tout $i \in \{1, \dots, m\}$ des combinaisons linéaires d'éléments de B et donc

$$\langle s_1, \dots, s_m \rangle_{\mathbb{L}} = \langle b_{1_1}, \dots, b_{1_{j_1}}, \dots, b_{m_1}, \dots, b_{m_{j_m}} \rangle_{\mathbb{L}}.$$

Le \mathbb{L} -module engendré par B est donc finiment engendré par des éléments de B lui-même.

Les mots k -automatiques admettent également une caractérisation en terme de leur k -noyau, qui s'ajoute aux caractérisations en termes d'arbre et d'AFDS. En effet, un mot $\mathbf{x} \in \mathbb{K}^{\mathbb{N}}$ est k -automatique si et seulement si $\ker_k(\mathbf{x})$ est fini. Les mots k -automatiques sont donc des mots k -réguliers.

Exemple 4.1.4. Soit \mathbf{t} le mot de Thue-Morse défini à l'Exemple 3.1.1. Le mot \mathbf{t} est 2-automatique, et peut être défini de la façon suivante pour tout $n \geq 0$

$$t_n = |\text{rep}_2(n)|_1 \bmod 2.$$

Les éléments de $\ker_2(\mathbf{t})$ étant de la forme $(t_{2^\ell n+r})_{n \geq 0}$ pour $\ell \geq 0$ et $r < 2^\ell$, nous avons pour tout $n \geq 0$

$$\text{rep}_2(2^\ell n + r) = \text{rep}_2(n)w,$$

où $\text{val}_2(w) = r$. Cela implique que si le nombre de 1 dans w est pair, $(t_{2^\ell n+r})_{n \geq 0} = \mathbf{t}$. Si il est impair, $(t_{2^\ell n+r})_{n \geq 0} = \bar{\mathbf{t}}$ où $\bar{0} = 1$ et $\bar{1} = 0$. On a donc

$$\ker_2(\mathbf{t}) = \{\mathbf{t}, \bar{\mathbf{t}}\}.$$

Le 2-noyau de \mathbf{t} est donc finiment engendré et même fini.

Considérons \mathcal{A} un alphabet. Une *série formelle* est une application $S : \mathcal{A}^* \rightarrow \mathbb{K}$ notée $S = \sum_{w \in \mathcal{A}^*} (S, w)w$ où $(S, w) = S(w)$ est le coefficient de w dans S . L'ensemble des séries formelles définies sur l'alphabet \mathcal{A} et à coefficients dans \mathbb{K} est noté $\mathbb{K}\langle\langle \mathcal{A} \rangle\rangle$. Une série $S \in \mathbb{K}\langle\langle \mathcal{A} \rangle\rangle$ est \mathbb{K} -reconnaissable s'il existe un entier $r > 0$, un morphisme $\mu : \mathcal{A}^* \rightarrow \mathbb{K}^{r \times r}$ et des matrices $\lambda \in \mathbb{K}^{1 \times r}$ et $\gamma \in \mathbb{K}^{r \times 1}$ telles que $\gamma\mu(w)\lambda = (S, w)$ pour tout $w \in \mathcal{A}^*$. Le triplet (λ, μ, γ) est alors appelé la *représentation linéaire* de S .

Soit $k \geq 2$ un entier. À toute suite $\mathbf{x} \in \mathbb{K}^{\mathbb{N}}$ peut être associée une série formelle $S_{\mathbf{x}}$ définie telle que $(S_{\mathbf{x}}, w) = x_{\text{val}_k(w)}$ pour tout $w \in \mathcal{A}_k^*$. Comme défini dans [5], une suite $\mathbf{x} \in \mathbb{K}^{\mathbb{N}}$ est k -régulière si la série formelle $S_{\mathbf{x}}$ est \mathbb{K} -reconnaissable.

Si on note le *produit de Hadamard* \odot et que l'on note également \underline{L} la série formelle associée au langage L , définie telle que $(\underline{L}, w) = 1$ si $w \in L$ et valant 0 sinon, alors la \mathbb{K} -reconnaissabilité de la série formelle $S_{\mathbf{x}}$ est équivalente à la \mathbb{K} -reconnaissabilité de la série formelle

$$G_{\mathbf{x}} := \sum_{w \in L_k} x_{\text{val}_k(w)}w = S_{\mathbf{x}} \odot \underline{L}_k. \quad (4.2)$$

grâce au résultat suivant.

Proposition 4.1.5. Soit $k \geq 2$ un entier. Le langage L_k étant reconnaissable, nous avons

- 1) la série \underline{L}_k est \mathbb{K} -reconnaissable,
- 2) si deux séries formelles $S, T \in \mathbb{K}\langle\langle \mathcal{A} \rangle\rangle$ sont \mathbb{K} -reconnaissables, alors $S \odot T$ est également \mathbb{K} -reconnaissable.

Les deux définitions des suites k -régulières sont évidemment équivalentes, car nous pouvons passer de l'une à l'autre grâce au Corollaire 2.3 de [5].

Exemple 4.1.6. Considérons la suite s_2 définie dans l'Exemple 4.1.4. À cette suite, nous pouvons associer la série formelle S_{s_2} définie comme

$$S_{s_2} = \sum_{w \in \mathcal{A}_2^*} s_2(\text{val}_2(w))w.$$

Nous devons trouver des matrices λ, γ et un morphisme μ telles que $\gamma\mu(w)\lambda = (S_{s_2}, w)$ pour tout $w \in \mathcal{A}_2^*$. Pour ce faire, nous définissons μ sur chacune des lettres de l'alphabet \mathcal{A}_2 ce qui nous permet de calculer $\mu(w) = \mu(w_0)\mu(w_1) \dots \mu(w_n)$. Nous définissons $\mu(0)$ et $\mu(1)$ de façon à ce que

$$\mu(w) = \begin{pmatrix} 1 & |w|_1 \\ 0 & 1 \end{pmatrix}$$

pour tout $w \in \mathcal{A}_2^*$, c'est-à-dire

$$\mu(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{et} \quad \mu(1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

De cette façon, les 0 dans w ne change rien au produit de matrices, et les 1 incrémentent le nombre en position $(1, 2)$. Par exemple,

$$\mu(11) = \mu(w) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}.$$

Si on définit $\gamma = \begin{pmatrix} 1 & 0 \end{pmatrix}$ et $\lambda = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, alors le produit $\gamma\mu(w)\lambda = |w|_1$, comme souhaité. La série S_{s_2} est donc \mathbb{N} -reconnaissable et s_2 est $(\mathbb{N}, 2)$ -régulière comme attendu au vu de l'Exemple 4.1.4.

4.2 Suites \mathcal{S} -régulières

Considérons un SNA $\mathcal{S} = (L, \mathcal{A}, <_{\mathcal{A}})$ construit sur un langage L clos par préfixe. Nous voulons définir les suites \mathcal{S} -régulière pour qu'elles gardent des propriétés similaires aux suites k -régulières et qu'elles généralisent les suites \mathcal{S} -automatiques. Pour ce faire, nous allons exploiter les deux définitions des suites k -régulières.

Nous allons commencer par généraliser la définition en terme de noyau. Soient $k \geq 2, \ell \geq 0, 0 \leq r \leq k^\ell$ et considérons un élément $(x_{k^\ell n+r})_{n \geq 0}$ du k -noyau de la

suite $\mathbf{x} \in \mathbb{K}^{\mathbb{N}}$. Parcourir les éléments de cette sous-suite dans $T_{\mathbf{x}}(L_k)$ correspond au fait de lire les décorations des noeuds atteints en lisant le même suffixe $w \in \mathcal{A}_k$ depuis chaque noeud de l'arbre, où w vérifie $\text{val}_k(w) = r$. En effet pour tout $n \geq 0$, le noeud décoré par $x_{k^\ell n+r}$ correspond au noeud atteint dans l'arbre en lisant le chemin labellisé par

$$\text{rep}_k(k^\ell n + r) = \text{rep}_k(n)w.$$

De cette façon, pour tout $n \geq 0$ le noeud atteint en lisant $\text{rep}_k(n)$ est x_n , duquel nous devons encore lire le même suffixe $w \in \mathcal{A}_k$. Chaque noeud distinct de la racine dans l'arbre $T(L_k)$ ayant k arêtes sortantes labellisées respectivement par les éléments de \mathcal{A}_k , la lecture du mot w peut se faire depuis n'importe quel noeud (et la décoration correspondante au noeud atteint existe toujours). Nous définissons donc le \mathcal{S} -noyau grâce à l'arbre $T_{\mathbf{x}}(L)$ de façon similaire, tout en ajoutant une restriction concernant les mots pouvant être lus depuis chaque noeud.

Définition 4.2.1. Soit $\mathcal{S} = (L, \mathcal{A}, <)$ un système de numération abstrait construit sur un langage clos par préfixe et régulier et soit $\mathbf{x} = x_0x_1 \cdots$ une suite de $\mathbb{K}^{\mathbb{N}}$. Pour tout mot u de \mathcal{A}^* , considérons la suite

$$\tau(\mathbf{x}, u) : n \mapsto \begin{cases} x_{\text{rep}_{\mathcal{S}}(n)u} & \text{si } \text{rep}_{\mathcal{S}}(n)u \in L \\ 0 & \text{sinon.} \end{cases}$$

Le \mathcal{S} -noyau de la suite \mathbf{x} est l'ensemble

$$\ker_{\mathcal{S}}(\mathbf{x}) := \{\tau(\mathbf{x}, u) \mid u \in \mathcal{A}^*\}.$$

Exemple 4.2.2. Considérons le SNA $\mathcal{S} = (a^*b^*, \{a, b\}, a < b)$ et la suite \mathbf{x} définie telle que $x_n = |\text{rep}_{\mathcal{S}}(n)|_a$ pour tout $n \geq 0$. L'arbre associé au langage a^*b^* et décoré par \mathbf{x} est représenté à la Figure 4.1. Le \mathcal{S} -noyau de la suite \mathbf{x} est composé des éléments

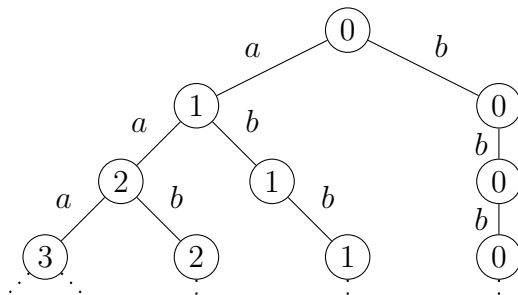


FIGURE 4.1 – Préfixe de l'arbre $T_{\mathbf{x}}(L)$ associé au langage $L = \{a^n b^m \mid n, m \in \mathbb{N}\}$.

repris dans la Table 4.1.

$u \in \{a, b\}^*$	$\tau(\mathbf{x}, u)$
ε	0,1,0,2,1,0,3,2,1,0 \cdots
a	1,2,0,3,0,0,4,0,0,0 \cdots
b	0,1,0,2,1,0,3,2,1,0 \cdots
aa	2,3,0,4,0,0,5,0,0,0 \cdots
ab	1,2,0,3,0,0,4,0,0,0 \cdots
b	0,1,0,2,1,0,3,2,1,0 \cdots
ba	0,0,0,0,0,0,0,0,0,0 \cdots
\vdots	

TABLE 4.1 – Premiers éléments de $\ker_{\mathcal{S}}(\mathbf{x})$.

La suite $\tau(\mathbf{x}, a)$ est obtenue en considérant les décorations des noeuds atteints en lisant a depuis chaque noeud de l'arbre. Or dans notre cas au vu de la signature labellisée de l'arbre $T(a^*b^*)$ donnée dans l'Exemple 1.3.14, la lettre a ne peut pas être lue depuis tous les noeuds. Si elle peut être lue, le noeud atteint a un a en plus dans sa représentation dans le système de numération abstrait \mathcal{S} . Ainsi, la suite $\tau(\mathbf{x}, a)$ est obtenue à partir de la suite \mathbf{x} , en ajoutant 1 aux indices de la suite correspondant aux noeuds depuis lesquels on peut lire la lettre a et en remplaçant les termes correspondant aux autres indices par 0. La suite $\tau(\mathbf{x}, ba) = (0)^\omega$ car le mot ba ne peut être lu de aucun noeud de $T(a^*b^*)$.

Les suites \mathcal{S} -régulières pour un SNA \mathcal{S} construit sur un langage clos par préfixe et régulier sont donc définies de la façon suivante.

Définition 4.2.3. Soit $\mathcal{S} = (L, \mathcal{A}, <)$ un système de numération abstrait construit sur un langage L clos par préfixe et régulier. Une suite \mathbf{x} à valeur dans \mathbb{K} est $(\mathbb{L}, \mathcal{S})$ -régulière si le \mathbb{L} -module engendré par le \mathcal{S} -noyau de \mathbf{x} est finiment engendré.

Remarque 4.2.4. Bien que les définitions soient similaires, il y a une différence entre le k -noyau d'une suite \mathbf{x} et le \mathcal{S}_k -noyau, où $\mathcal{S}_k = (L_k, \mathcal{A}_k, <)$. En effet, si u est un mot de \mathcal{A}_k^* ayant 0 pour préfixe, alors $\text{rep}_{\mathcal{S}_k}(0)u = \varepsilon u = u$ qui n'appartient pas à L_k par définition. Cela implique donc que la suite $\tau(\mathbf{x}, u)$ commence par un 0, tandis que la suite $(x_{k|u|n+\text{val}_k(u)})_{n \geq 0}$ commence par $x_{\text{val}_k(u)}$ qui est potentiellement différent de 0. Ces deux sous-suites analogues de \mathbf{x} ne sont donc pas nécessairement égales sur leur premier terme.

La Remarque 4.2.4 ne pose cependant pas problème car la suite \mathbf{x} sera k -régulière si et seulement elle est $(\mathbb{L}, \mathcal{S}_k)$ -régulière. (Voir [8]).

Exemple 4.2.5. Soit $\mathcal{S} = (a^*b^*, \{a, b\}, a < b)$ et considérons la suite \mathbf{x} telle que $x_n = |\text{rep}_{\mathcal{S}}(n)|_a$. Les premiers éléments de $\ker_{\mathcal{S}}(\mathbf{x})$ sont repris dans la Table 4.1. Nous

allons essayer de trouver des relations linéaires entre les éléments du \mathcal{S} -noyau de \mathbf{x} . Au vu des observations déjà réalisées dans l'Exemple 4.2.2, nous savons que pour tous $n, m > 0$ entiers, $\tau(\mathbf{x}, b^n a^m) = (0)^\omega$ car $b^n a^m$ ne peut être lu depuis aucun noeud de l'arbre $T(a^* b^*)$. De plus, pour tout $n > 0$ entier, $\tau(\mathbf{x}, b^n) = \mathbf{x}$ car la lecture de b dans l'arbre depuis n'importe quel noeud l mène à un noeud k ayant la même décoration (car $\text{rep}_{\mathcal{S}}(k) = \text{rep}_{\mathcal{S}}(l)b$ et donc $|\text{rep}_{\mathcal{S}}(k)|_a = |\text{rep}_{\mathcal{S}}(l)|_a$). Si on note \mathbf{y} la suite caractéristique de $\tau(\mathbf{x}, a)$, alors pour tout $n > 0$ entier la suite $\tau(\mathbf{x}, a^n) = \tau(\mathbf{x}, a) + n \cdot \mathbf{y}$. Tous les éléments du \mathcal{S} -noyau peuvent donc être réexprimés en termes des suites $\mathbf{x}, \tau(\mathbf{x}, a)$ et \mathbf{y} . La suite \mathbf{x} est donc $(\mathbb{N}, \mathcal{S})$ -régulière.

Nous allons à présent généraliser la définition de la k -régularité en terme de \mathbb{K} -reconnaissabilité de séries formelles, en utilisant la définition suivante issue de [8].

Définition 4.2.6. Soit $\mathcal{S} = (L, \mathcal{A}, <_{\mathcal{A}})$ un SNA. Une suite $\mathbf{x} \in \mathbb{K}^{\mathbb{N}}$ est $(\mathbb{K}, \mathcal{S})$ -régulière si la série formelle

$$G_{\mathbf{x}, \mathcal{S}} := \sum_{w \in L} x_{\text{val}_{\mathcal{S}}(w)} w \quad (4.3)$$

est \mathbb{K} -reconnaissable.

La Définition 4.2.6 généralise celle en terme de la série formelle (4.2), car dans la série formelle (4.3) nous sommes sur les mots de L qui vérifie $L = \{\text{rep}_{\mathcal{S}}(n) \mid n \in \mathbb{N}\}$.

Exemple 4.2.7. Continuons l'Exemple 4.2.5 et montrons que la suite \mathbf{x} est aussi $(\mathbb{N}, \mathcal{S})$ -régulière selon la Définition 4.2.6. La suite $G_{\mathbf{x}, \mathcal{S}}$ est dans ce cas de la forme suivante

$$\begin{aligned} G_{\mathbf{x}, \mathcal{S}} &= \sum_{w \in L} x_{\text{val}_{\mathcal{S}}(w)} w \\ &= \sum_{w \in L} |\text{rep}_{\mathcal{S}}(\text{val}_{\mathcal{S}}(w))|_a w \\ &= \sum_{w \in L} |w|_a w \end{aligned}$$

Nous devons donc définir des matrices γ, λ et un morphisme μ tels que pour tout $w \in a^* b^*$ nous avons $\gamma \mu(w) \lambda = |w|_a$. En suivant le raisonnement de l'Exemple 4.1.6 en remplaçant l'alphabet $\{0, 1\}$ par $\{a, b\}$, nous obtenons une représentation linéaire de $G_{\mathbf{x}, \mathcal{S}}$ donnée par

$$\mu(a) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{et} \quad \mu(b) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

et par

$$\gamma = \begin{pmatrix} 1 & 0 \end{pmatrix} \quad \text{et} \quad \lambda = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Ces matrices sont telles que $\gamma\mu(w)\lambda = |w|_a$ pour tout $w \in L$ pour les mêmes raisons que dans l'Exemple 4.1.6. La suite \mathbf{x} est donc $(\mathbb{N}, \mathcal{S})$ -régulière car elle admet (γ, μ, λ) comme représentation linéaire.

Une nouvelle fois, grâce au Théorème 29 de [8], les définitions de la $(\mathcal{S}, \mathbb{K})$ -régularité en terme de noyau (Définition 4.2.3) et en terme de série formelle (Définition 4.2.6) sont équivalentes et illustrées dans les Exemples 4.2.5 et 4.2.7.

4.2.1 Caractérisation des suites \mathcal{S} -régulières en terme d'arbre

Nous allons dans cette section introduire une nouvelle caractérisation de la \mathcal{S} -régularité pour un SNA construit sur un langage clos par préfixe. Pour ce faire nous allons définir de nouveaux arbres appelés *arbres h -linéaires*. Soit T un arbre décoré et $h \geq 1$. Sommairement, nous dirons que T est h -linéaire si les décorations du niveau h de ses sous-arbres de hauteur h peuvent être obtenues par combinaison linéaire des décorations des noeuds du niveau 0 à $h - 1$.

Définition 4.2.8. Soit L un langage clos par préfixe sur un alphabet \mathcal{A} fini. Soit $h \geq 1$ et notons $r(h) \geq 1$ le nombre de facteurs labellisés de hauteur h deux à deux distincts de l'arbre $T(L)$. Nous avons donc $F_h = \{T_1, \dots, T_{r(h)}\}$. L'arbre décoré $T_{\mathbf{x}}(L)$ est (\mathbb{L}, h) -linéaire (ou simplement h -linéaire) si, pour tout $1 \leq j \leq r(h)$ et tout mot $w \in L$ tels que $T[w, h] = T_j$, il existe pour tout mot u appartenant à $\text{dom}(T_j) \cap \mathcal{A}^h$ des constantes $c_{j,u,v} \in \mathbb{L}$ telles que

$$x_{wu} = \sum_{\substack{v \in \text{dom}(T_j) \\ |v| < h}} c_{j,u,v} x_{wv}.$$

Par la suite pour tout $w \in L$ tels que le sous-arbre de hauteur h $T[w, h]$ de $T(L)$ est égal à T_j pour $1 \leq j \leq r(h)$, nous disons que $T[w, h]$ est de type T_j .

Exemple 4.2.9. Nous allons utiliser dans cet exemple le système de numération positionnel U associé à la suite de Fibonacci, défini à l'Exemple 2.3.2. Pour rappel, le langage $L_U = 1\{0, 01\}^* \cup \{\varepsilon\}$. L'arbre labellisé $T(L_U)$ associé au langage L_U a donc la spécificité suivante :

- Si un noeud est atteint par une arête labellisée par 0, alors le chiffre 0 et le chiffre 1 peuvent être lus à partir de ce noeud. Il a donc deux arêtes sortantes labellisées respectivement par 0 et 1.
- Si un noeud est atteint par une arête labellisée par 1, seul un 0 peut être lu après. Il a donc une unique arête sortante labellisée par 0.

Les facteurs non isomorphes (sans les décorations) de $T(L_U)$ de hauteur 2 sont représentés à la Figure 4.2.

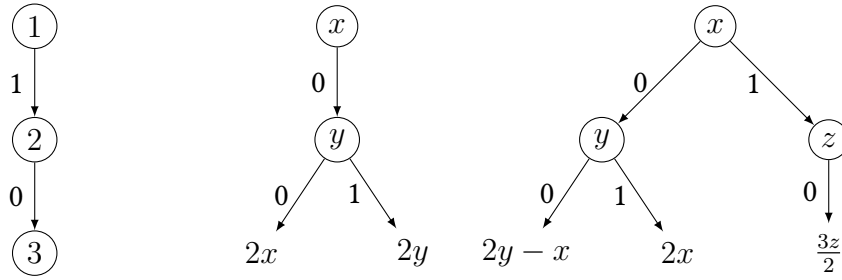


FIGURE 4.2 – Facteur de $T(F)$ de hauteur 2.

Pour rendre l'arbre $T(L_U)$ 2-linéaire, il faut commencer par décorer les trois premiers niveaux de l'arbre. Ensuite, pour continuer de décorer les noeuds de l'arbre, il faut donner les relations linéaires entre les deux premiers niveaux et le troisième niveau des facteurs non isomorphes de $T(L_U)$. Avec les relations linéaires et les décorations du préfixe de hauteur 2 de $T(L_U)$ donné à la Figure 4.2, nous pouvons construire l'arbre décoré dont le préfixe de hauteur 5 est représenté dans la Figure 4.3. Le parcours en largeur de noeuds de l'arbre $T(L_U)$ décoré, nous donne la suite \mathbf{x} suivante

$$\mathbf{x} = 1, 2, 3, 4, 4, 5, 6, 6, 6, 8, 9, 8, 8, \dots$$

comptant le nombre d'entrées distinctes de 0 dans la n -ième ligne du triangle de Pascal généralisé basé sur les représentations de Fibonacci (voir [14] pour plus de détails).

La nouvelle notion d'arbre (\mathbb{L}, h) -linéaire nous permet d'établir une caractérisation des suites (\mathbb{L}, k) -régulières, qui est généralisable aux suites $(\mathbb{L}, \mathcal{S})$ -régulières pour un SNA construit sur un langage clos par préfixe.

Théorème 4.2.10. *Soit $k \geq 2$ un entier. Une suite $\mathbf{x} = x_0x_1 \dots$ à valeur dans \mathbb{K} est (\mathbb{L}, k) -régulière si et seulement si il existe $h \geq 1$ tel que l'arbre décoré $T_{\mathbf{x}}(L_k)$ est (\mathbb{L}, h) -linéaire.*

Démonstration. Condition nécessaire. Si le \mathbb{L} -module M engendré par le k -noyau est finiment engendré, alors par la Remarque 4.1.3 nous pouvons supposer qu'il existe un nombre fini d'élément de $\ker_k(\mathbf{x})$ engendrant $\langle \ker_k(\mathbf{x}) \rangle_{\mathbb{L}}$. Ainsi, il existe $h \geq 1$ tel que pour tout $0 \leq r < k^h$, la suite $(x_{k^h n+r})_{n \geq 0}$ est une combinaison linéaire des suites de la forme $(x_{k^j n+s})_{n \geq 0}$ pour $0 \leq j < h$ et $0 \leq s < k^j$.¹ De façon explicite, il existe des constantes $c_{r,j,s}$ telles que

$$\forall n \geq 0, \quad x_{k^h n+r} = \sum_{j=0}^{h-1} \sum_{s=0}^{k^j-1} c_{r,j,s} x_{k^j n+s}. \quad (4.4)$$

1. Nous aurons donc $\langle \ker_k(\mathbf{x}) \rangle_{\mathbb{L}} = \langle \{(x_{k^j n+s})_{n \geq 0} \mid 0 \leq j < h, 0 \leq s < k^j\} \rangle_{\mathbb{L}}$

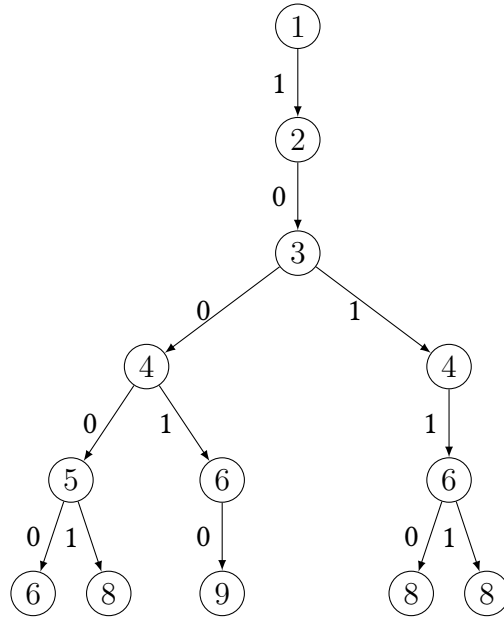


FIGURE 4.3 – Premiers niveaux de $T(L_U)$ décoré.

Cela étant, il nous reste à montrer que les décorations des noeuds de niveau h des sous-arbres de hauteur h de $T_{\mathbf{x}}(L_k)$ peuvent s'écrire comme une combinaison linéaire des décorations des noeuds des $h - 1$ autres niveaux. Pour tout $w \in L_k$, nous avons l'égalité suivante

$$\text{val}_k(w \cdot \text{dom}(T[w, h])) = \{k^j \text{val}_k(w) + s \mid 0 \leq j \leq h, 0 \leq s < k^j\}.$$

Notons que lorsque $j = h$, dans le membre de droite, cela signifie que le mot $u \in \text{dom}(T[w, h])$ vérifie $|u| = h$ et que le noeud $x_{k^h \text{val}_k(w) + \text{val}_k(u)}$ est sur le niveau h du sous-arbre $T[w, h]$. L'Egalité (4.4) étant vérifiée pour tout $n \geq 0$ (et donc en particulier pour $n = \text{val}_k(w)$ pour tout $w \in L_k$), cela signifie que l'arbre $T(L_k)$ est (\mathbb{L}, h) -linéaire et que pour tout sous-arbre $T[w, h]$ les décorations des noeuds du dernier niveau sont obtenues par l'Egalité (4.4) comme combinaison linéaire des décorations des noeuds des autres niveaux.

Condition suffisante. Supposons que l'arbre $T_{\mathbf{x}}(L_k)$ est (\mathbb{L}, h) -linéaire. Cela implique que pour chaque sous-arbre non décoré non isomorphe de $T_{\mathbf{x}}(L_k)$ de hauteur h , il existe une relation linéaire du type de (4.4) liant les décorations des noeuds. Les deux types de facteurs de hauteur h de $T_{\mathbf{x}}(L_k)$ sont son préfixe de hauteur h et les facteurs $T[w, h]$ pour $w \neq \varepsilon$.² Ainsi, pour $0 \leq r < k^h$, la suite $(x_{k^h n + r})_{n \geq 1}$ est combinaison linéaire des suites $(x_{k^j n + s})_{n \geq 1}$ pour $0 \leq j < h$ et $0 \leq s < k^j$. La suite

2. Cela est dû au fait que l'arbre $T_{\mathbf{x}}(L_k)$ n'a pas de boucle 0 sur sa racine car les mots de L_k ne

$(x_{k^h n+r})_{n \geq 0}$ est donc combinaison linéaire des suites $(x_{k^j n+s})_{n \geq 0}$ pour $0 \leq j < h$ et $0 \leq s < k^j$ et de la suite \mathbf{z} définie telle que $z_i = 0$ pour tout $i \geq 1$. Pour montrer que la suite \mathbf{x} est k -régulière, il faut encore montrer que pour tout $m \geq h$ et pour tout $0 \leq \ell < k^m$, les suites $(x_{k^m n+\ell})_{n \geq 0}$ sont des combinaisons linéaires des suites $(x_{k^j n+s})_{n \geq 1}$ pour $0 \leq j < h$ et $0 \leq s < k^j$ et de \mathbf{z} .
 Pour $t \geq 1$ et $0 \leq p < k^{h+t}$, si nous écrivons $p = k^h q_p + r_p$, nous avons

$$(x_{k^{h+t} n+p})_{n \geq 1} = (x_{k^h(k^t n+q_p)+r_p})_{n \geq 1}.$$

Cette suite peut donc se réécrire comme une combinaison linéaire des suites de la forme $(x_{k^j(k^t n+q_p)+s})_{n \geq 1}$ pour $0 \leq j < h$ et $0 \leq s < k^j$. Si $j+t \geq h$, nous pouvons réécrire $j+t = h+t'$ et recommencer l'étape précédente jusqu'à avoir une combinaison linéaire des suites $(x_{k^j n+s})_{n \geq 1}$ pour $0 \leq j < h$ et $0 \leq s < k^j$. La suite $(x_{k^{h+t} n+p})_{n \geq 0}$ est alors une combinaison linéaire des suites $(x_{k^j n+s})_{n \geq 1}$ pour $0 \leq j < h$ et $0 \leq s < k^j$ et de \mathbf{z} . \square

Contrairement à l'arbre $T(L_k)$ étudié dans le cas de la k -régularité, dans lequel tous les sous-arbres sont isomorphes sans décoration, la lecture des mots de L dans l'arbre $T(L)$ ne mène pas nécessairement à des sous-arbres de hauteur h isomorphes. Nous allons donc affiner la définition du \mathcal{S} -noyau en faisant intervenir le type de l'arbre atteint en lisant $\text{rep}_{\mathcal{S}}(n)$ dans l'arbre associé au langage L .

Définition 4.2.11. Soient $h \geq 1$, $\mathbf{x} \in \mathbb{K}^{\mathbb{N}}$ et $\mathcal{S} = (L, \mathcal{A}, <)$ un système de numération abstrait construit sur un langage L régulier et clos par préfixe. Le nombre $r(h)$ de sous-arbres de hauteur h non décorés de $T(L)$ est fini et sont notés $T_1, \dots, T_{r(h)}$. Pour tout $u \in \mathcal{A}^*$ et tout $j \in \{1, \dots, r(h)\}$, considérons la suite

$$\tau(\mathbf{x}, u, T_j) : n \mapsto \begin{cases} x_{\text{rep}_{\mathcal{S}}(n)u} & \text{si } \text{rep}_{\mathcal{S}}(n)u \in L \text{ et } T[\text{rep}_{\mathcal{S}}(n), h] = T_j \\ 0 & \text{sinon.} \end{cases}$$

Le \mathcal{S} -noyau h -filtré de la suite \mathbf{x} est l'ensemble

$$\ker_{h,\mathcal{S}}(\mathbf{x}) := \{\tau(\mathbf{x}, u, T_j) \mid u \in \mathcal{A}^*, 1 \leq j \leq r(h)\}.$$

Pour tout $1 \leq j \leq r(h)$, la suite caractéristique χ_j est définie de la façon suivante

$$\chi_j : n \mapsto \begin{cases} 1 & \text{si } T[\text{rep}_{\mathcal{S}}(n), h] = T_j \\ 0 & \text{sinon} \end{cases}$$

Par définition du produit de Hadamard \odot nous obtenons

$$\tau(\mathbf{x}, u) \odot \chi_j = \tau(\mathbf{x}, u, T_j).$$

commencent pas par 0.

En particulier, nous pouvons remarquer le \mathcal{S} -noyau 0-filtré de \mathbf{x} est simplement le \mathcal{S} -noyau et que pour tout $u \in A^*$

$$\tau(\mathbf{x}, u) = \sum_{j=1}^{r(h)} \tau(\mathbf{x}, u, T_j). \quad (4.5)$$

Exemple 4.2.12. Reprenons l'Exemple 4.2.2. L'arbre $T(a^*b^*)$ possède deux sous-arbres distincts T_1 et T_2 de hauteur 2 représentés à la Figure 4.4. Le sous-arbre de la forme

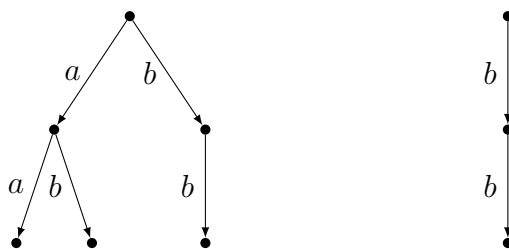


FIGURE 4.4 – Types de sous arbres de l'arbre $T(a^*b^*)$.

T_1 n'est atteint que lorsque la dernière lettre du mot lu dans l'arbre $T(a^*b^*)$ est a . Par exemple, la sous-suite $\tau(\mathbf{x}, b, T_1)$ est donc

$$2, 4, 0, 7, 0, 0, 11, \dots$$

Les occurrences de 0 dans cette sous-suite sont les mêmes que dans la sous-suite $\tau(\mathbf{x}, a)$ car les sous-arbres de type T_1 sont les seuls dans lesquels on peut lire a depuis la racine. La suite $\tau(\mathbf{x}, b, T_2)$ est donnée par

$$0, 0, 5, 0, 8, 9, 0, \dots$$

Au vu de l'expression de la suite $\tau(\mathbf{x}, b)$ dans la Table 4.1, nous avons bien $\tau(\mathbf{x}, b) = \tau(\mathbf{x}, b, T_1) + \tau(\mathbf{x}, b, T_2)$.

Le lemme suivant nous en apprend plus sur le lien présent entre ces deux noyaux.

Lemme 4.2.13. *Le \mathbb{L} -module $\langle \ker_{\mathcal{S}}(\mathbf{x}) \rangle_{\mathbb{L}}$ engendré par le \mathcal{S} -noyau de \mathbf{x} est finiment engendré si et seulement si le \mathbb{L} -module $\langle \ker_{h,\mathcal{S}}(\mathbf{x}) \rangle_{\mathbb{L}}$ engendré par le \mathcal{S} -noyau h -filtré de \mathbf{x} est finiment engendré pour tout $h \geq 0$.*

Démonstration. Condition nécessaire. Supposons que $\langle \ker_{\mathcal{S}}(\mathbf{x}) \rangle_{\mathbb{L}}$ est finiment engendré. Par la Remarque 4.1.3, nous pouvons supposer que $\langle \ker_{\mathcal{S}}(\mathbf{x}) \rangle_{\mathbb{L}}$ est engendré par des éléments $s_1, \dots, s_m \in \ker_{\mathcal{S}}(\mathbf{x})$. Les éléments \mathbf{m} de $\ker_{h,\mathcal{S}}(\mathbf{x})$ peuvent se réécrire comme $\mathbf{y} \odot \chi_j$ pour $\mathbf{y} \in \ker_{\mathcal{S}}(\mathbf{x})$ et pour $0 \leq j \leq r(h)$, où les χ_j sont définis à la

Définition 4.2.11. Au vu du fait que \mathbf{y} peut se réécrire comme une combinaison linéaire de s_1, \dots, s_m , cela implique que \mathbf{m} peut se réécrire comme une combinaison linéaire des éléments $s_1 \odot \chi_j, \dots, s_m \odot \chi_j$. Le \mathcal{S} -noyau h -filtré de \mathbf{x} est donc finiment engendré par les suites $s_i \odot \chi_j$ pour $i \in \{1, \dots, m\}$ et pour $0 \leq j \leq r(h)$.

Condition suffisante. Au vu de la l'égalité 4.5, nous pouvons déduire que $\langle \ker_{\mathcal{S}}(\mathbf{x}) \rangle_{\mathbb{L}}$ est inclus dans $\langle \ker_{h,\mathcal{S}}(\mathbf{x}) \rangle_{\mathbb{L}}$. Tout sous-module d'un module finiment engendré sur un anneau Noethérien est également finiment engendré. Cela implique que si $\langle \ker_{h,\mathcal{S}}(\mathbf{x}) \rangle_{\mathbb{L}}$ est finiment engendré, alors $\langle \ker_{\mathcal{S}}(\mathbf{x}) \rangle_{\mathbb{L}}$ l'est également. \square

Nous allons maintenant démontrer le théorème liant (\mathbb{L}, h) -linéarité de l'arbre $T_{\mathbf{x}}(L)$ et la $(\mathbb{L}, \mathcal{S})$ -régularité de \mathbf{x} . Commençons par montrer que la régularité de la suite implique qu'il existe h pour lequel l'arbre est linéaire.

Proposition 4.2.14. *Soit $\mathcal{S} = (L, \mathcal{A}, <)$ un système de numération abstrait construit sur un langage L clos par préfixe (mais pas nécessairement régulier). Si le \mathbb{L} -module $\langle \ker_{\mathcal{S}}(\mathbf{x}) \rangle_{\mathbb{L}}$ engendré par le \mathcal{S} -noyau de la suite \mathbf{x} à valeurs dans \mathbb{K} est finiment engendré, alors il existe $h \geq 1$ tel que l'arbre décoré $T_{\mathbf{x}}(L)$ est (\mathbb{L}, h) -linéaire.*

Démonstration. En utilisant la Remarque 4.1.3, nous pouvons supposer que le noyau $\langle \ker_{\mathcal{S}}(\mathbf{x}) \rangle_{\mathbb{L}}$ est engendré par un nombre fini de suites données par $\tau(\mathbf{x}, u_1), \dots, \tau(\mathbf{x}, u_m)$. Soit $h = 1 + \max_i |u_i|$. Comme nous l'avons montré dans le lemme précédent, le noyau $\langle \ker_{h,\mathcal{S}}(\mathbf{x}) \rangle_{\mathbb{L}}$ est finiment engendré par des éléments de la forme $\tau(\mathbf{x}, u_i) \odot \chi_j = \tau(\mathbf{x}, u_i, T_j)$ pour $i \in \{1, \dots, m\}$ et $j \in \{1, \dots, r(h)\}$. Soit v un mot de longueur h et soit $j \in \{1, \dots, r(h)\}$ tel que $v \in \text{dom}(T_j)$. La suite $\tau(\mathbf{x}, v, T_j)$ étant un élément de $\langle \ker_{h,\mathcal{S}}(\mathbf{x}) \rangle_{\mathbb{L}}$, c'est une combinaison linéaire d'éléments de la forme $\tau(\mathbf{x}, u_i, T_p)$ pour $i \in \{1, \dots, m\}$ et $p \in \{1, \dots, r(h)\}$. Cependant, au vu de la forme de la suite $\tau(\mathbf{x}, v, T_j)$, sa décomposition n'utilise que des éléments de la forme $\tau(\mathbf{x}, u_i, T_j)$. Pour tout n tel que $\text{rep}_{\mathcal{S}}(n)$ est la racine d'un arbre de type T_j , cela signifie que la décoration $x_{\text{rep}_{\mathcal{S}}(n)v}$ est une combinaison linéaire à coefficients dans \mathbb{L} des décorations $x_{\text{rep}_{\mathcal{S}}(n)u_i}$. Au vu du fait que $|v| = h$, cela signifie que les décorations des feuilles de n'importe quel facteur de hauteur h est combinaison linéaire à coefficients dans \mathbb{L} des décorations des noeuds sur les niveaux de 0 à $h - 1$ (car $|u_i| \leq h - 1$). L'arbre $T_{\mathbf{x}}(L)$ est donc (\mathbb{L}, h) -linéaire. \square

Afin de passer à la réciproque, i.e. montrer que la linéarité de l'arbre implique la régularité de la suite, nous avons besoin de deux lemmes intermédiaires.

Lemme 4.2.15. *Si un arbre décoré $T_{\mathbf{x}}(L)$ est h -linéaire, alors il est $(h + 1)$ -linéaire.*

Démonstration. Soit $T[w, h + 1]$ un sous-arbre de hauteur $h + 1$ de $T_{\mathbf{x}}(L)$ et supposons que wa_1, \dots, wa_ℓ sont les fils du noeud w . Les noeuds wa_1, \dots, wa_ℓ sont donc les racines des sous-arbres $T[wa_1, h], \dots, T[wa_\ell, h]$ de $T[w, h + 1]$ de hauteur h et disjoints

(sans noeud en commun). L'arbre $T_{\mathbf{x}}(L)$ étant h -linéaire, nous savons que pour tout $0 \leq j \leq \ell$ les décorations des noeuds de niveau h de $T[wa_j, h]$ sont combinaisons linéaires des décorations des noeuds de $T[wa_j, h - 1]$. Les noeuds des $T[wa_j, h - 1]$ étant des noeuds de $T[w, h]$, cela implique que les noeuds de niveau $h + 1$ de $T[w, h + 1]$ sont combinaisons linéaires des noeuds de $T[w, h]$ et que $T_{\mathbf{x}}(L)$ est $(h + 1)$ -linéaire. \square

Lemme 4.2.16. *Si L est un langage régulier et clos par préfixe, alors la fonction qui compte le nombre de sous-arbres de hauteur n de $T(L)$ est bornée par le nombre d'états de l'automate minimal émondé³ acceptant L .*

Démonstration. De par la définition de l'automate minimal dans la Section 1.2.1, le fait que le langage $L \subseteq \mathcal{A}^*$ soit fini implique que l'automate \mathbb{A}_L a un nombre fini d'états noté N . Chacun des états de l'automate minimal sont distingués par définition, i.e. pour toutes paires d'états $\{q, q'\}$ de \mathbb{A}_L il existe un mot $w \in \mathcal{A}^*$ tel que $\delta_L(q, w) \in F$ et $\delta_L(q', w) \notin F$ ou inversement.⁴ Au vu du nombre d'états de \mathbb{A}_L , les états sont distingués par un mot $u \in \mathcal{A}^*$ de longueur $|u| \leq N - 1$. En termes d'arbres, cela peut se reformuler en disant que si la lecture d'un mot w (respectivement w') dans l'automate minimal mène à un état q (respectivement q'), alors il existe un mot u de longueur $h \leq N - 1$ tel que u appartient à $\text{dom}(T[w, h])$ et pas à $\text{dom}(T[w', h])$ ou inversement. \square

Ces lemmes suffisent pour montrer la condition inverse.

Proposition 4.2.17. *Soit $\mathcal{S} = (L, \mathcal{A}, <)$ un système de numération abstrait construit sur un langage L régulier et clos par préfixe. S'il existe $h \geq 1$ tel que l'arbre $T_{\mathbf{x}}(L)$ décoré par la suite \mathbf{x} à valeurs dans \mathbb{K} est (\mathbb{L}, h) -linéaire, alors la suite \mathbf{x} est $(\mathbb{L}, \mathcal{S})$ -régulière.*

Démonstration. Par le Lemme 4.2.15, nous pouvons choisir h de façon à ce que les facteurs de $T(L)$ soient en bijection avec les états de l'automate minimal de L et respectant la fonction de transition de \mathbb{A}_L , i.e. si un état q correspond à un facteur $T[w, h]$, alors $q \cdot a$ correspond à $T[wa, h]$. Nous pouvons donc dénoter les sous-arbres de hauteur h de $T_{\mathbf{x}}(L)$ par leur racine T_w à la place de $T_1, \dots, T_{r(h)}$ (nous avons $T_w = T'_w$ si w et w' correspondent au même état de \mathbb{A}_L).

Par hypothèse, il existe $h \geq 1$ tel que l'arbre $T_{\mathbf{x}}(L)$ est (\mathbb{L}, h) -linéaire. Plus précisément, cela signifie que si w est la racine d'un sous-arbre $T[w, h]$ de hauteur h , alors pour tout mot u de longueur h tels que $wu \in L$, il existe des constantes $c_{w,u,v} \in \mathbb{L}$ telles que

$$x_{wu} = \sum_{\substack{v \in \text{dom}(T_w) \\ |v| < h}} c_{w,u,v} x_{wv}. \quad (4.6)$$

3. Cela signifie que les états puits ont été retirés.

4. En effet, si on considère deux états de \mathbb{A}_L qui sont de la forme $u^{-1}L$ et $v^{-1}L$ respectivement, il existe un mot $w \in \mathcal{A}^*$ tel que $wu \in L$ et $wv \notin L$ ou inversement (sinon on aurait $u^{-1}L = v^{-1}L$ et les deux états ne seraient pas distincts).

Si u est tel que $wu \notin L$, nous poserons que les constantes valent 0. De cette façon, $\tau(\mathbf{x}, u, T_w)$ est une combinaison linéaire à coefficients dans \mathbb{L} de suites de la forme $\tau(\mathbf{x}, v, T_w)$ ⁵. Nous allons montrer que l'élément $u' = au$ vérifiant $|u'| = h+1$ et $wau \in L$, peut se réécrire comme une combinaison linéaire à coefficients dans \mathbb{L} d'éléments de la forme $\tau(\mathbf{x}, v, T_w)$ pour $|v| < h$. et En utilisant l'Egalité 4.6 dans la première et la dernière égalité, nous avons

$$\begin{aligned}
x_{wau} &= \sum_{\substack{v \in \text{dom}(T_{wa}) \\ |v| < h}} c_{wa,u,v} x_{wav} \\
&= \sum_{\substack{v \in \text{dom}(T_{wa}) \\ |v| < h-1}} c_{wa,u,v} x_{wav} + \sum_{\substack{v \in \text{dom}(T_{wa}) \\ |v| = h-1}} c_{wa,u,v} x_{wav} \\
&= \sum_{\substack{v \in \text{dom}(T_{wa}) \\ |v| < h-1}} c_{wa,u,v} x_{wav} + \sum_{\substack{v \in \text{dom}(T_{wa}) \\ |v| = h-1}} c_{wa,u,v} \sum_{\substack{z \in \text{dom}(T_w) \\ |z| < h}} c_{w,av,z} x_{wz}.
\end{aligned}$$

Cela signifie que $\tau(\mathbf{x}, au, T_w)$ est une combinaison linéaire à coefficients dans \mathbb{L} de suites de la forme $\tau(\mathbf{x}, av, T_w)$ avec $|av| < h$ et de suites de la forme $\tau(\mathbf{x}, z, T_w)$ avec $|z| < h$.

Le même raisonnement peut s'appliquer pour des mots u'' tels que $|u'| < |u''|$ en répétant l'argument précédent, ce qui permet de montrer que tout élément de $\ker_{h,S}(\mathbf{x})$ peut se réécrire comme une combinaison linéaire à coefficients dans \mathbb{L} d'éléments de la forme $\tau(\mathbf{x}, v, T_w)$ pour $|v| < h$ et $w \in A^*$. Nous pouvons donc conclure par le Lemme 4.2.13. \square

Les deux Propositions 4.2.14 et 4.2.17 se résument en le théorème suivant.

Théorème 4.2.18. *Soit $\mathcal{S} = (L, A, <)$ un système de numération abstrait construit sur un langage clos par préfixe et régulier. La suite \mathbf{x} à valeurs dans \mathbb{K} est une suite $(\mathbb{L}, \mathcal{S})$ -régulière si et seulement si il existe $h \geq 1$ tel que l'arbre décoré $T_{\mathbf{x}}(L)$ est (\mathbb{L}, h) -linéaire.*

4.3 Suites $\frac{p}{q}$ -régulières

Le Théorème 4.2.18 que nous venons d'obtenir nous donne une caractérisation des suites \mathcal{S} -régulières pour les SNA construits sur un langage régulier et clos par préfixe. Pour tous p, q tels que $p > q > 1$, le langage $L_{\frac{p}{q}}$ n'est pas régulier et le SNA associé n'entre donc pas dans les conditions d'application du Théorème 4.2.18.

5. Remarquons que ici nous réexprisons des éléments du $\ker_{h,S}(\mathbf{x})$ comme combinaison linéaire d'autres éléments de $\ker_{h,S}(\mathbf{x})$

Tout comme la suite s_2 est 2-régulière, il serait souhaitable que la définition de la régularité dans le cas des SNA construits sur un langage clos par préfixe mais non régulier soit telle que la généralisation à la base $\frac{3}{2}$ de la suite s_2 soit $\frac{3}{2}$ -régulière. Considérons le SNA $\mathcal{S} = (L_{\frac{3}{2}}, \mathcal{A}_3, <)$ et la suite $(s(n))_{n \geq 0}$ définie dans l'Exemple 3.1.11. (L'arbre $T_{s(n)}(L_{\frac{3}{2}})$ est représenté à la Figure 4.5).

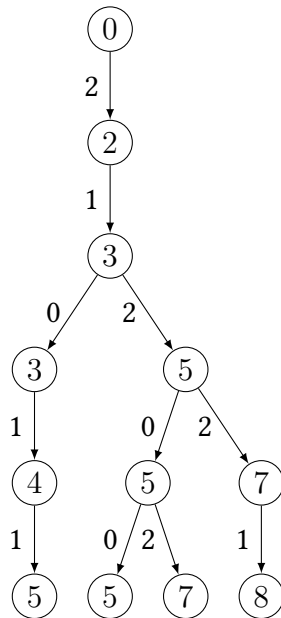


FIGURE 4.5 – Premiers niveaux de $T(L_{\frac{3}{2}})$ décoré par $(s(n))_{n \geq 0}$.

Afin d'obtenir une définition cohérente de la $\frac{p}{q}$ -régularité d'une suite, nous allons essayer de voir si un des noyaux associés à la suite est finiment engendré (i.e. s'il existe des relations linéaires entre les éléments du noyau). En utilisant le package Mathematica "IntegerSequence"⁶ permettant de tester la k -régularité d'une suite pour k entier, nous voyons que dans le cas de la suite $s(n)$, le 3-noyau, 2-noyau et le 6-noyau n'ont pas de relations linéaires entre leurs éléments et la suite n'est donc pas 2-régulière, 3-régulière ou 6-régulière.

Pour ce qui est du \mathcal{S} -noyau associé au SNA $\mathcal{S} = (L_{\frac{3}{2}}, \mathcal{A}_3, <)$, les premiers éléments sont donnés dans la Table 4.2.

Tous les éléments du \mathcal{S} -noyau correspondant à des mots u de longueur 1 ou 2 ne sont pas repris dans la Table 4.2 car certains de ces éléments peuvent se réexprimer en fonction d'autres éléments. En effet, les suites $\tau(s(n), 0)$ et $\tau(s(n), 2)$ ont des 0 aux

6. La description du package Mathematica est disponible sur <https://ericrowland.github.io/packages.html>

$u \in \{0, 1, 2\}^*$	$\tau(s(n), u)$
ε	0,2,3,3,5,4,5,7,5,5,7,8,...
0	0,0,3,0,5,0,5,0,5,0,7,0,5,0,6...
1	0,3,0,4,0,5,0,8,0,6,0,9,0,8...
10	0,3,0,0,0,5,0,0,0,6,0,0,8...
01	0,0,4,0,0,0,6,0,0,0,8,0,0,7...
11	0,0,0,5,0,0,0,9,0,0,0,10,0,0...
00	0,0,0,0,5,0,0,0,5,0,0,0,5,0...
000	0,0,0,0,0,0,0,0,5,0,0,0,0,0...

TABLE 4.2 – Premiers éléments du $\ker_S(s(n))$.

positions impairs car au vu de la signature $(02, 1)^\omega$ de l'arbre $T(L_{\frac{3}{2}})$, les chiffres 0 et 2 ne peuvent être lus que depuis un noeud sur 2. Pour un $n \geq 0$ entier fixé, le n -ième terme de $\tau(s(n), 0)$ et de $\tau(s(n), 2)$ sont la somme des chiffres de $\text{rep}_{\frac{3}{2}}(n)$ à laquelle on additionne respectivement 0 ou 2. Nous pouvons donc passer de la suite $\tau(\mathbf{x}, 0)$ à la suite $\tau(\mathbf{x}, 2)$ en ajoutant 0 aux positions impaires et 2 aux positions paires. Cela revient donc à écrire $\tau(s(n), 2) = \tau(s(n), 0) + (20)^\omega$.

Au vu du Lemme 3.2.6, pour tout $h \geq 1$ l'arbre $T(L_{\frac{3}{2}})$ possède 2^h sous-arbres de hauteur h de domaine distinct (en ne comptant pas le préfixe de hauteur h). De ce fait, si un mot u de longueur 2 peut être lu dans un sous-arbre de hauteur 2 donné, ça ne sera le cas dans aucun des 3 autres (au vu du fait qu'il y en a 4 en tout et que leurs domaines sont disjoints). De cette façon, dans la suite $\tau(s(n), u)$ les termes non nuls sont séparés par $2^{|u|} - 1$ termes qui valent 0. Pour les mots u tels que $|u| = j$, les éléments du \mathcal{S} -noyau peuvent avoir leurs termes non nuls en des positions r modulo 2^j , pour r vérifiant $0 \leq r < 2^j$. Les suites caractéristiques des ensembles $\mathbb{N}, 2\mathbb{N}, 2\mathbb{N}+1, \dots, 2^j\mathbb{N}, 2^j\mathbb{N}+1, \dots, 2^j\mathbb{N}+2^j-1$ étant linéairement indépendantes, le \mathcal{S} -noyau de $s(n)$ ne peut pas être finiment engendré.

Aucun des noyaux associés à la suite $s(n)$ définis à la Définition 4.2.3 et en (4.1) ne semblent avoir de relations linéaires convenables entre ses éléments. Ainsi la $\frac{p}{q}$ -régularité va être définie en terme de la h -linéarité de l'arbre $T(L_{\frac{p}{q}})$. Au vu des Théorèmes 4.2.10 et 4.2.18, cette définition est bien une généralisation des définitions de la régularité introduites précédemment.

Définition 4.3.1. Soient p et q des entiers premiers entre eux tels que $p > q > 1$. Une suite \mathbf{x} à valeurs dans \mathbb{K} est $(\mathbb{L}, \frac{p}{q})$ -régulière lorsqu'il existe $h \geq 1$ tel que l'arbre décoré $T_{\mathbf{x}}(L_{\frac{p}{q}})$ est (\mathbb{L}, h) -linéaire.

Remarque 4.3.2. La suites $s(n)$ est bien $\frac{3}{2}$ -régulière selon la Définition 4.3.1, comme

montré dans l'Exemple 4.3.5.

4.3.1 Propriétés classiques conservées

Nous allons décrire dans cette sous section des propriétés classiques pour les suites k -régulières qui sont conservées dans le cas des suites $\frac{p}{q}$ -régulières.

Nous notons dans la suite (w_0, \dots, w_{q-1}) la signature de l'arbre $T(L_{\frac{p}{q}})$.

Proposition 4.3.3. *Si une suite \mathbf{x} est $(\mathbb{L}, \frac{p}{q})$ -régulière et prend un nombre fini de valeurs, alors elle est $\frac{p}{q}$ -automatique.*

Démonstration. Supposons que \mathbf{x} est une suite $\frac{p}{q}$ -régulière et prend un nombre fini de valeurs. Par définition, cela implique qu'il existe $h \geq 1$ tel que $T(L_{\frac{p}{q}})$ est (\mathbb{L}, h) -linéaire. Cela signifie que pour tout facteur de F_h^∞ , les décorations des feuilles sont combinaisons linéaires des décorations des noeuds des $h - 1$ niveaux précédents. Tout élément de F_{h-1}^∞ ne produit donc qu'au plus un facteur de $F_{h,w_j,0}^\infty$ pour tout $j \in \{0, \dots, q-1\}$. De ce fait pour tout $0 \leq j \leq q - 1$, nous avons $F_{h-1,w_j,0}^\infty \leq F_h^\infty$ et la suite \mathbf{x} est donc $\frac{p}{q}$ -automatique par la Proposition 3.2.13. \square

Par définition des suites k -automatiques en terme de leur noyau, nous savons que leur k -noyau est fini, et donc en particulier finiment engendré, ce qui implique que les suites k -automatiques sont donc également k -régulières lorsque k est un entier $k \geq 2$. Les mots $\frac{p}{q}$ -automatiques quant à eux ne sont pas nécessairement $\frac{p}{q}$ -réguliers. Considérons le mot infini $\mathbf{x} = q_0q_1q_213155133322 \dots$ engendré par l'AFDS représenté à la Figure 4.6.

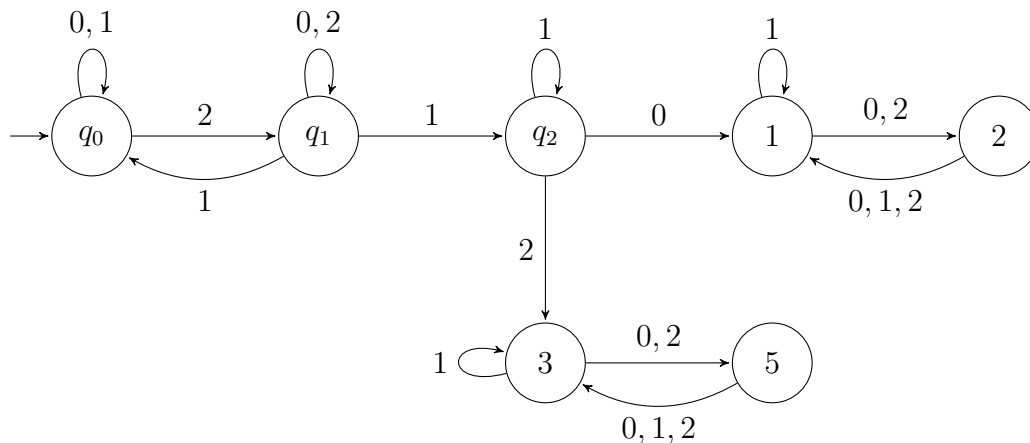


FIGURE 4.6 – L'AFDS engendrant le mot \mathbf{x} .

Les nombres 1,2 et 3,5 ont été choisis de façon à ce que les vecteurs $(1, 2)$ et $(3, 5)$ soient linéairement indépendants. Au vu du fait que \mathbf{x} est engendré par un AFDS prenant en argument des mots de $L_{\frac{3}{2}}$, il est $\frac{3}{2}$ -automatique. Nous allons montrer qu'il n'est pas $\frac{3}{2}$ -régulier par l'absurde, en supposant qu'il existe $h \geq 1$ tel que $T_{\mathbf{x}}(L_{\frac{3}{2}})$ est h -linéaire. Soit $u = 1^h 0$ un mot de longueur $h+1$ sur l'alphabet \mathcal{A}_3 . Grâce au Corollaire 3.2.8, nous savons que u est suffixe d'un mot de $L_{\frac{3}{2}}$ et qu'il existe donc un mot $w_0 \in \mathcal{A}_3^*$ tel que $w_0 u \in L_{\frac{3}{2}}$. Soit $R \in \{0, \dots, 2^h - 1\}$ l'entier vérifiant $\text{val}_{\frac{3}{2}}(w_0) \equiv R \pmod{2^{h+1}}$. Pour tout $w \in L_{\frac{3}{2}}$ vérifiant $\text{val}_{\frac{3}{2}}(w) \equiv R \pmod{2^{h+1}}$, nous savons par la Proposition 3.2.6 que le sous-arbre de hauteur $h+1$ atteint en lisant w est isomorphe (sans les décorations) à celui atteint en lisant w_0 . En particulier, pour tout $k \geq 0$ entier le mot $w_k \in L_{\frac{3}{2}}$ représentant l'entier $\text{val}_{\frac{3}{2}}(w_0) + k \cdot 2^{h+1}$ vérifie $w_k u \in L_{\frac{3}{2}}$. Pour tout $n \geq 0$ entier, notons

$$g_0(n) = \#\{x \in \mathcal{A}_3^n \mid 210x \in L_{\frac{3}{2}}\} \text{ et } g_2(n) = \#\{x \in \mathcal{A}_3^n \mid 212x \in L_{\frac{3}{2}}\}.$$

Au vu de la forme de l'arbre $T(L_{\frac{3}{2}})$, les mots de $L_{\frac{3}{2}}$ de longueur supérieure à 3 commencent soit par le préfixe 210 ou le préfixe 212, ce qui implique que pour tout $n \geq 0$ entier, $g_0(n) + g_2(n)$ est le nombre de mots de longueur $n+3$. Dans l'arbre $T(L_{\frac{3}{2}})$, cela signifie que le nombre de noeuds au niveau $n+3$ est $g_0(n) + g_2(n)$. Pour tout $n \geq 2$ les suites $g_0(n)$ et $g_2(n)$ sont strictement croissantes, et il existe donc un entier $N \geq 2$ (que nous supposons pair sans perte de généralité) tel que $g_0(N) > 2^{h+1}$ et $g_2(N) > 2^{h+1}$. En d'autres mots, le nombre de noeuds au niveau $N+3$ de l'arbre $T(L_{\frac{3}{2}})$ est strictement supérieur à $2^{h+1} + 2^{h+1}$, comme représenté à la Figure 4.7.

Au vu du fait que le niveau $N+3$ est composé de plus de 2^{h+2} noeuds, dont les 2^{h+1} premiers sont atteints en lisant des mots ayant pour préfixe 210, il existe un mot w_i de la forme $210x$ tel que $w_i u \in L_{\frac{3}{2}}$ et qui est de longueur impaire (N est pair, donc $N+3$ est impair). Le même raisonnement est d'application pour les 2^{h+1} derniers noeuds du niveau $N+3$, qui sont atteints en lisant des mots ayant pour préfixe 212. Il existe donc un mot w_j de la forme $212x'$, tel que $w_j u \in L_{\frac{3}{2}}$ et de longueur impaire. Pour tout $h \geq 1$, il n'existe qu'un unique sous-arbre de hauteur $h+1$ de $T(L_{\frac{3}{2}})$ dans lequel le mot u peut être lu. Pour $h=4$, ce sous-arbre est représenté à la Figure 4.8.

Au vu du choix de w_i et de w_j , leur lecture dans l'arbre mène à des sous-arbres de hauteur $h+1$ qui sont isomorphes (sans les décorations). La situation est illustrée à la Figure 4.9.

Les sous-arbres décorés atteints lors de la lecture de w_i et w_j dans l'arbre sont représentés à la Figure 4.8. Si l'arbre $T(L_{\frac{3}{2}})$ est h -linéaire, cela signifie que dans les sous-arbres de hauteur $h+1$ atteints lors de la lecture de w_i, w_j , il existe une relation linéaire entre les noeuds internes des h derniers niveaux et la première feuille qui doit être identique dans les deux sous-arbres. Plus précisément il doit exister des coefficients

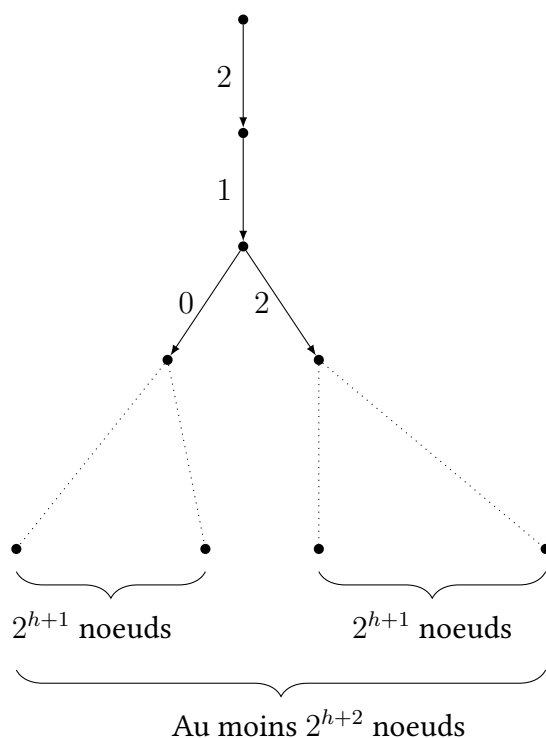


FIGURE 4.7 – Préfixe de hauteur $N + 3$ de $T(L_{\frac{3}{2}})$.

c_0, c_1, \dots, c_{h-1} tels que

$$2 = c_0.1 + c_1.1 + \dots + c_{h-1}.1 \quad \text{et} \quad 5 = c_0.3 + c_1.3 + \dots + c_{h-1}.3$$

Cela est impossible car $2.3 \neq 5$, ce qui nous permet de conclure.

D'autres propriétés des suites régulières en lien avec les suites automatiques sont conservées avec la définition des suites $\frac{p}{q}$ -régulières.

Proposition 4.3.4. *Soit $m \geq 2$ un entier. Si une suite $\mathbf{x} \in \mathbb{Z}^{\mathbb{N}}$ est $(\mathbb{Z}, \frac{p}{q})$ -régulière, alors la suite $\mathbf{y} = (x_n \bmod m)_{n \geq 0}$ est $\frac{p}{q}$ -automatique.*

Démonstration. S'il existe $h \geq 1$ tel que la suite \mathbf{x} est (\mathbb{Z}, h) -régulière, cela signifie que l'arbre $T_{\mathbf{x}}(L_{\frac{p}{q}})$ est (\mathbb{Z}, h) -linéaire. En particulier, par le Lemme 4.2.15, l'arbre $T_{\mathbf{x}}(L_{\frac{p}{q}})$ est également $(\mathbb{Z}, h + 1)$ -linéaire. Grâce à la Proposition 3.2.6, nous savons que le nombre de facteurs de hauteur h de $T(L_{\frac{p}{q}})$ est fini. Au vu du fait que la suite $\mathbf{y} = (x_n \bmod m)_{n \geq 0}$ prend un nombre fini de valeurs parmi $\{0, \dots, m - 1\}$, cela signifie que le nombre de facteurs de hauteur h décorés de $T_{\mathbf{y}}(L_{\frac{p}{q}})$ est fini. Si on considère un élément de F_h^{∞} , on sait par $(\mathbb{Z}, h + 1)$ -linéarité de $T_{\mathbf{x}}(L_{\frac{p}{q}})$ que cet élément s'étend

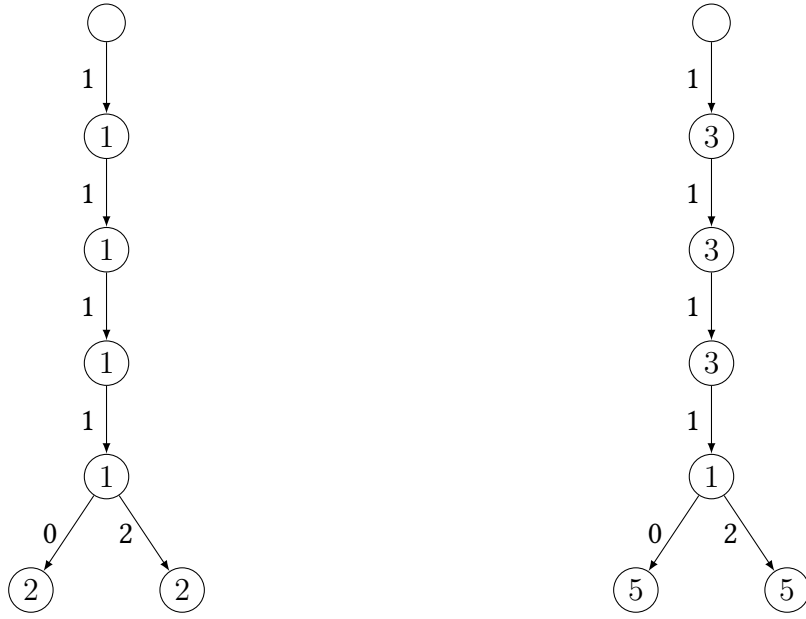


FIGURE 4.8 – Sous arbres de $T(L_{\frac{3}{2}})$ atteints en lisant w_i et w_j pour $h = 4$.

en un unique élément de $F_{h,w_j,0}^\infty$. En effet, l'arbre $T_{\mathbf{x}}(L_{\frac{p}{q}})$ étant $h + 1$ -linéaire, nous savons que les décorations des feuilles d'un élément de $F_{h+1,w_j,0}^\infty$ sont uniquement déterminées par une relation linéaire entre les décorations des h premiers niveaux. Ces relations linéaires étant conservées modulo m , nous pouvons à nouveau conclure par la Proposition 3.2.13. \square

4.3.2 Comment montrer qu'un arbre est (\mathbb{L}, h) -linéaire?

La définition de la $\frac{p}{q}$ -régularité d'une suite \mathbf{x} reposant sur la notion de h -linéarité de l'arbre $T_{\mathbf{x}}(L_{\frac{p}{q}})$, il pourrait être intéressant d'établir un protocole permettant d'essayer de répondre à la question "Existe-t-il $h \geq 1$ tel que l'arbre $T_{\mathbf{x}}(L_{\frac{p}{q}})$ est h -linéaire?".

Exemple 4.3.5. Soient $\frac{p}{q} = \frac{3}{2}$. Nous allons tester la $\frac{3}{2}$ -linéarité de la suite $(s(n))_{n \geq 0}$ en utilisant la Définition 4.3.1. Nous savons par la Proposition 3.2.6 que l'arbre $T(L_{\frac{3}{2}})$ a 4 facteurs de hauteur 2 non isomorphes en tant que sous-arbres non décorés. De plus, si deux sous arbres de hauteur 2 sont isomorphes, alors les mots w et w' de $L_{\frac{3}{2}}$ menant à leur racine vérifient $\text{val}_{\frac{3}{2}}(w) \equiv \text{val}_{\frac{3}{2}}(w') \pmod{4}$. Le sous-arbre correspondant à $\text{val}_{\frac{3}{2}}(w) \equiv 2 \pmod{4}$ est le sous-arbre de hauteur 2 atteint en lisant la rep $\frac{3}{2}(2)$ dans l'arbre $T(L_{\frac{3}{2}})$ et est représenté à la Figure 4.10. Il a été décoré avec les décorations correspondant à sa première occurrence dans l'arbre $T_{s(n)}(L_{\frac{3}{2}})$. Nous allons es-

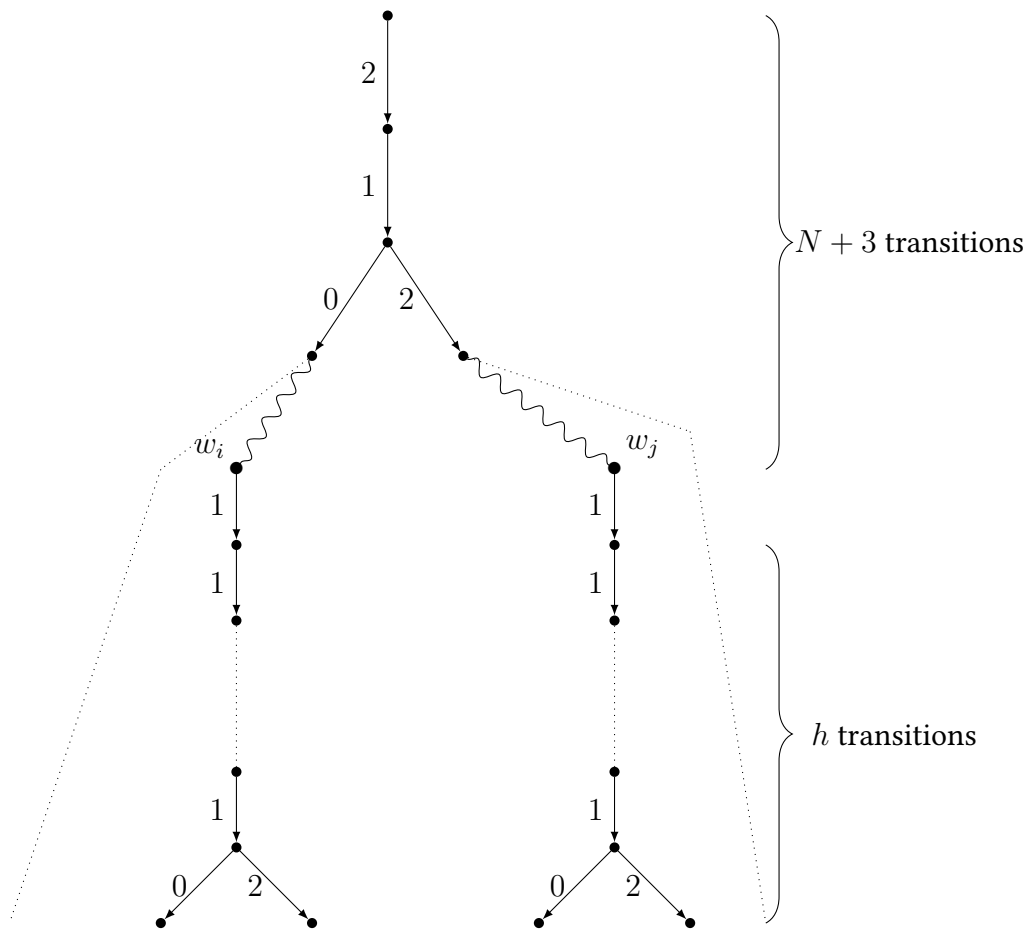


FIGURE 4.9 – Préfixe de $T(L_{\frac{3}{2}})$.

sayer de conjecturer des relations linéaires pour chacune des feuilles de ce sous-arbre particulier. Afin d'obtenir le système permettant d'essayer de deviner les relations linéaires entre les noeuds internes et les feuilles, nous allons utiliser les 4 premières occurrences de ce sous-arbre particulier dans l'arbre (ce sous-arbre ayant 3 noeuds internes, il faut considérer au moins 3 occurrences du sous-arbres dans $T_{s(n)}(L_{\frac{3}{2}})$ pour avoir un nombre d'équations supérieur ou égal au nombre d'inconnues.

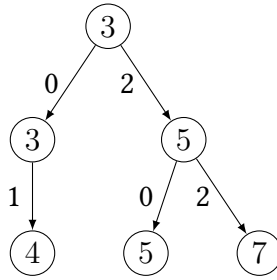


FIGURE 4.10 – Sous-arbre de hauteur 2 ayant pour racine $\text{rep}_{\frac{3}{2}}(2)$ dans $T_{s(n)}(L_{\frac{3}{2}})$.

La Table 4.3 contient sur chacune de ses lignes les décorations des noeuds d'une des 4 premières occurrences des sous-arbres. Les 3 premières colonnes sont les décorations des noeuds internes, et les 3 dernières sont les décorations des feuilles.

$x_{n_j,1}$	$x_{n_j,2}$	$x_{n_j,3}$	$x_{n_j,4}$	$x_{n_j,5}$	$x_{n_j,6}$
3	3	5	4	5	7
7	7	9	8	9	11
5	5	7	6	7	9
6	6	8	7	8	10

TABLE 4.3 – Décorations des noeuds des 8 premières occurrence du sous arbre considéré.

Pour chacune des feuilles de ce sous-arbre, nous pouvons maintenant établir un système de 4 équations à 3 inconnues. Si l'arbre $T_{s(n)}(L_{\frac{3}{2}})$ est 3-linéaire, les relations linéaires liant les noeuds internes à la première feuille dans ce sous-arbre doivent être les mêmes dans les 3 autres sous-arbres isomorphes considérés. Le système correspondant à la première feuille peut être réécrit de la façon suivante,

$$\begin{pmatrix} 3 & 3 & 5 \\ 7 & 7 & 9 \\ 5 & 5 & 7 \\ 6 & 6 & 8 \end{pmatrix} \begin{pmatrix} \alpha_{2,1,1} \\ \alpha_{2,1,2} \\ \alpha_{2,1,3} \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \\ 6 \\ 7 \end{pmatrix}.$$

Une solution de ce système est donnée par $\alpha_{2,1,1} = 0$, $\alpha_{2,1,2} = \frac{1}{2}$ et $\alpha_{2,1,3} = \frac{1}{2}$. En résolvant un système similaire pour chacune des feuilles du sous-arbre considéré, nous obtenons les relations linéaires représentées à la Figure 4.11.

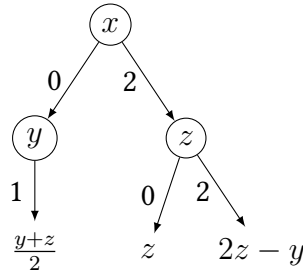


FIGURE 4.11 – Sous-arbre de hauteur 2 ayant pour racine $w \in L_{\frac{3}{2}}$ vérifiant $\text{val}_{\frac{3}{2}}(w) \equiv 2 \pmod{4}$ de $T_{s(n)}(L_{\frac{3}{2}})$ avec les relations linéaires entre ses décorations.

D'un point de vue logique, les décorations des feuilles s'expliquent facilement. Au vu de la décoration x de la racine, la première feuille devrait être décorée de $x + 1$ et nous avons bien $\frac{y+z}{2} = \frac{x+0+x+2}{2} = x + 1$. Un raisonnement similaire est possible pour les autres feuilles.

Les systèmes pour obtenir les relations linéaires entre les décorations des feuilles et des noeuds des sous-arbres correspondant à $R \in \{0, 1, 3\}$ sont similaires, et les 3 types de sous-arbres restant sont représentés à la Figure 4.12.

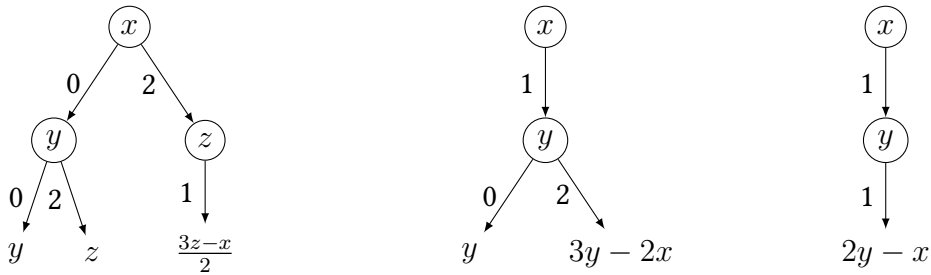


FIGURE 4.12 – Sous-arbres de hauteur 2 ayant pour racine $w \in L_{\frac{3}{2}}$ vérifiant $\text{val}_{\frac{3}{2}}(w) \equiv 0 \pmod{4}$, $\text{val}_{\frac{3}{2}}(w) \equiv 1 \pmod{4}$ ou $\text{val}_{\frac{3}{2}}(w) \equiv 3 \pmod{4}$ de $T_{s(n)}(L_{\frac{3}{2}})$ avec les relations linéaires entre ses décorations.

Il existe donc un $h \geq 1$ pour lequel l'arbre $T(L_{\frac{3}{2}})$ est (\mathbb{Q}, h) -linéaire et la suite $s(n)$ est donc $(\mathbb{Q}, \frac{3}{2})$ -régulière par définition.

Nous allons maintenant décrire une procédure générale permettant de tester la h -linéarité de l'arbre $T_{\mathbf{x}}(L_{\frac{p}{q}})$, pour \mathbf{x} un mot infini. Supposons que nous connaissions les M premiers termes d'une suite \mathbf{x} , et que cela nous permet de décorer les ℓ premiers niveaux de l'arbre $T(L_{\frac{p}{q}})$. Soit $h \geq 1$. Nous voudrions essayer de trouver des relations

linéaires valables liant les décorations des feuilles et les décorations des $h - 1$ premiers niveaux des sous-arbres non isomorphes de hauteur h .

Par la Proposition 3.2.6, nous savons qu'il y a q^h facteurs non décorés de hauteur h de l'arbre $T(L_{\frac{p}{q}})$ et distincts du préfixe de hauteur h . Deux facteurs $T[u_j, h]$ et $T[u_i, h]$ sont isomorphes si $\text{val}_{\frac{p}{q}}(u_j) \equiv \text{val}_{\frac{p}{q}}(u_i) \pmod{q^h}$. Le nombre de feuilles de ces q^h facteurs étant p^h , il y a p^h relations linéaires à déterminer. Soit $R \in \{0, \dots, q^h - 1\}$ et considérons l'ensemble fini des sous-arbres $T[u_j, h]$, $j \in J_R$ tels que $\text{val}_{\frac{p}{q}}(u_j) \equiv R \pmod{q^h}$. Tous les arbres $T[u_j, h]$ pour $j \in J_R$ étant isomorphes, notons k_R leur nombre de feuilles et i_R leur nombre de noeuds internes (i.e. les noeuds qui ne sont pas des feuilles). En parcourant en largeur les noeuds des sous-arbres du type $T[u_j, h]$ pour $j \in J_R$, nous pouvons pour tout $j \in J_R$ énumérer leur décoration,

$$\underbrace{x_{n_j,1}, \dots, x_{n_j,i_R}}_{\text{Les noeuds internes}}, \underbrace{x_{n_j,i_R+1}, \dots, x_{n_j,i_R+k_R}}_{\text{Les feuilles}}$$

Pour tout $j \in J_R$ et pour $t \in \{1, \dots, k_R\}$, considérons l'équation

$$\alpha_{R,t,1}x_{n_j,1} + \dots + \alpha_{R,t,i_R}x_{n_j,i_R} = x_{n_j,i_R+t}$$

où les coefficients $\alpha_{R,t,i}$ ne dépendent pas de j car deux feuilles correspondantes dans deux facteurs isomorphes de hauteur h doivent vérifier la même relation linéaire. Pour j fixé dans J_R , nous avons donc un ensemble k_R équations correspondant aux k_R feuilles de l'arbre $T[u_j, h]$.

Pour j variant dans J_R , nous obtenons un système de $\#J_R$ équations à i_R inconnues $\alpha_{R,t,1}, \dots, \alpha_{R,t,i_R}$. Si $\#J_R \geq i_R$ (i.e. le nombre d'équations est plus grand que le nombre d'inconnues) et que le système n'a pas de solution, l'hypothèse de h -linéarité de l'arbre peut alors être rejetée et l'hypothèse de $(h + 1)$ -linéarité peut être testée. Si le système a une solution, nous pouvons conjecturer que les coefficients de la relation linéaire entre les noeuds internes et la feuille $i_R + t$ de chaque facteur du type $T[u_j, h]$ est donnée par cette solution.

Il reste maintenant à répéter la même opération pour déterminer les coefficients de la relation linéaire liant les noeuds internes et les autres feuilles, et ce dans les q^h facteurs non isomorphes de hauteur h .

Table des figures

1.1	Un premier AFDS.	7
1.2	Un arbre pointé ordonné de racine v_0	10
1.3	Arbre lexicographique associé au langage $L = \{a^n b^m \mid n, m \in \mathbb{N}\}$	11
1.4	Sous arbre \mathcal{T}_L , $\mathcal{T}_L[b]$ et $\mathcal{T}_L[ba]$ de l'arbre lexicographique associé à $L = \{a^n b^m \mid n, m \in \mathbb{N}\}$	12
1.5	Arbre $T(L)$ associé au langage $L = \{a^n b^m \mid n, m \in \mathbb{N}\}$	14
1.6	Arbre associé à la signature périodique labellisée $\mathbf{s} = (02, 1)^\omega$	16
3.1	L'AFDS engendrant le mot de Thue-Morse.	29
3.2	L'AFD associé à la substitution T	33
3.3	L'AFDS engendrant le mot de Thue-Morse généralisé.	34
3.4	Arbre associé à la signature labellisée purement périodique $\mathbf{s} = \left(0 \cdots (l_0 - 1), l_0 \cdots l_0 + l_1 - 1, \dots, \sum_{j < r-1} l_j \cdots \left(\left(\sum_{j < r} l_j\right) - 1\right)\right)^\omega$	39
3.5	Illustration des fils du noeud x_m	40
3.6	Arbre associé à la signature labellisée purement périodique $(01 \cdots (k - 1))^\omega$	42
3.7	Premiers niveaux de l'arbre $T(\mathbf{s})$ associé à la signature labellisée purement périodique $\mathbf{s} = (w_0, \dots, w_{r-1})$	43
3.8	Arbre associé au morphisme σ_i	43
3.9	Illustration de l'Equation (3.4).	45
3.10	Unique facteur de hauteur 1 permettant de construire un facteur de hauteur 2.	47
3.11	Extensions des facteurs de hauteur 1 en facteurs de hauteur 2 dans l'arbre $T(L_{\frac{3}{2}})$	54
3.12	Transitions entre deux arbres de F_h^∞	57
3.13	Les deux extensions de l'arbre U	59
4.1	Préfixe de l'arbre $T_{\mathbf{x}}(L)$ associé au langage $L = \{a^n b^m \mid n, m \in \mathbb{N}\}$	64
4.2	Facteur de $T(F)$ de hauteur 2.	68
4.3	Premiers niveaux de $T(L_U)$ décoré.	69

4.4	Types de sous arbres de l'arbre $T(a^*b^*)$	71
4.5	Premiers niveaux de $T(L_{\frac{3}{2}})$ décoré par $(s(n))_{n \geq 0}$	75
4.6	L'AFDS engendrant le mot \mathbf{x}	77
4.7	Préfixe de hauteur $N + 3$ de $T(L_{\frac{3}{2}})$	79
4.8	Sous arbres de $T(L_{\frac{3}{2}})$ atteints en lisant w_i et w_j pour $h = 4$	80
4.9	Préfixe de $T(L_{\frac{3}{2}})$	81
4.10	Sous-arbre de hauteur 2 ayant pour racine $\text{rep}_{\frac{3}{2}}(2)$ dans $T_{s(n)}(L_{\frac{3}{2}})$	82
4.11	Sous-arbre de hauteur 2 ayant pour racine $w \in L_{\frac{3}{2}}$ vérifiant $\text{val}_{\frac{3}{2}}(w) \equiv 2 \pmod{4}$ de $T_{s(n)}(L_{\frac{3}{2}})$ avec les relations linéaires entre ses décorations.	83
4.12	Sous-arbres de hauteur 2 ayant pour racine $w \in L_{\frac{3}{2}}$ vérifiant $\text{val}_{\frac{3}{2}}(w) \equiv 0 \pmod{4}$, $\text{val}_{\frac{3}{2}}(w) \equiv 1 \pmod{4}$ ou $\text{val}_{\frac{3}{2}}(w) \equiv 3 \pmod{4}$ de $T_{s(n)}(L_{\frac{3}{2}})$ avec les relations linéaires entre ses décorations.	83

Bibliographie

- [1] AKIYAMA, Shigeki, Christiane FROUGNY et Jacques SAKAROVITCH. Powers of rationals modulo 1 and rational base number systems. *Israel Journal of Mathematics*. 2008, 168 (1), p. 53-91.
- [2] ALLOUCHE, Jean-Paul et Jeffrey SHALLIT. *Automatic Sequences : Theory, Applications, Generalizations*. Cambridge University Press, 2003.
- [3] ALLOUCHE, Jean-Paul et Jeffrey SHALLIT. The ring of k -regular sequences. *Theoret. Comput. Sci.* 1992, 98, p. 163-197.
- [4] BERSTEL, Jean, Luc BOASSON, Olivier CARTON et Isabelle FAGNOT. Sturmian trees. *Theory of Computing Systems*. 2009, 46 (3), p. 443-478.
- [5] BERSTEL, Jean et Christophe REUTENAUER. *Noncommutative Rational Series with Applications*. Cambridge University Press, 2010.
- [6] BRUYÈRE, Véronique et Georges HANSEL. Bertrand numeration systems and recognizability. *Theoretical Computer Science*. 1997, 181 (1), p. 17-43.
- [7] CARPI, A., A. de LUCA et S. VARRICCHIO. Special factors and uniqueness conditions in rational trees. *Theory of Computing Systems*. 2001, 34 (4), p. 375-395.
- [8] CHARLIER, Emilie, Célia CISTERNINO et Manon STIPULANTI. Regular sequences and synchronized sequences in abstract numeration systems. *European Journal of Combinatorics*. 2022, 101.
- [9] COBHAM, Alan. On the hartmanis-stearns problem for a class of tag machines. *In : 9th Annual Symposium on Switching and Automata Theory*, IEEE Computer Society, 1968, p. 51-60.
- [10] COBHAM, Alan. Uniform tag sequences. *Mathematical Systems Theory*. 1972, 6 (1-2), p. 164-192.
- [11] CONRAD, Keith. *Noetherian Modules online*(visité le 21. août 2023). Disp. à l'adr. <<https://kconrad.math.uconn.edu/blurbs/linmultialg/noetherianmod.pdf>>.
- [12] DEKKING, Michel. *The Thue-Morse sequence in base 3/2*. arXiv, 2023. Disp. à l'adr. DOI : 10.48550/ARXIV.2301.13563.

- [13] KOLAKOSKI, William. Elementary problem 5304. *The American Mathematical Monthly*. 1965, 72, p. 674.
- [14] LEROY, Julien, Michel RIGO et Manon STIPULANTI. Counting the number of non-zero coefficients in rows of generalized pascal triangles. *Discrete Mathematics*. 2017, 340 (5), p. 862-881.
- [15] MAES, Arnaud. *Morphic predicates and applications to the decidability of arithmetic theories*. Thèse de doct. Université de Mons-Hainaut. 1999.
- [16] MARSAULT, Victor. *Énumération et numération*. Thèse de doct. Telecom Paris. 2015.
- [17] MCCULLOCH, Warren S. et Walter PITTS. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*. 1943, 5 (4), p. 115-133.
- [18] OLDENBURGER, Rufus. Exponent trajectories in symbolic dynamics. *Trans. Amer. Math. Soc.* 1939, 46 (0), p. 453-466.
- [19] RABBIN, Michael O. et Dana S. SCOTT. Finite automata and their decision problem. *IBM J. Res. Dev.* 1959, 3, p. 114-125.
- [20] RIGO, Michel. *Abstract numeration systems on a regular language and recognizability*. Thèse de doct. Université de Liège. 2001.
- [21] RIGO, Michel. *Formal Languages, Automata and Numeration Systems : Applications to Recognizability and Decidability*. T. 2. Wiley & Sons, Incorporated, John, 2014.
- [22] RIGO, Michel. Generalization of automatic sequences for numeration systems on a regular language. *Theoretical Computer Science*. 2000, 244, p. 271-281.
- [23] RIGO, Michel. *Théorie des automates et langages formels*. 2009–2010. Note de cours de l'université de Liège.
- [24] RIGO, Michel et Arnaud MAES. More on generalized automatic sequences. *Journal of Automata, Languages, and Combinatorics*. 2002, 7, p. 351-376.
- [25] RIGO, Michel et Manon STIPULANTI. Automatic sequences : from rational bases to trees. *Discrete mathematics and theoretical computer science*. 2021, 24 (1), p. 25.
- [26] RIGO, Michel et Manon STIPULANTI. Revisiting regular sequences in light of rational base numeration systems. *Discrete Mathematics*. 2022, 345 (3), p. 112735.
- [27] THUE, Axel. *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*. J. Dybwad, Kristiania, 1912.
- [28] TURING, Alan. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*. 1937, s2-42 (1), p. 230-265.